Vol. 5, No. 1 (1979)

Edited by
George J Klir

International Journal of

GENERAL SYSTEMS

Methodology • Applications • Education

UNIVERSITY OF HAWAI
LIBRARY
APR 12 11 02 AM '79

PERIODICAL
Q 295
I 58
LIBRARY

Gordon and Breach Science Publishers
New York London Paris

International Journal of

GENERAL SYSTEMS

A Comprehensive Periodical devoted to General Systems Methodology, Applications and Education

EDITOR George J. Klir, *Department of Systems Science, School of Advanced Technology, State University of New York, Binghamton, New York 13901, U.S.A.*

EDITORIAL BOARD

Carl F. Albrecht, *Department of Pharmacology, University of Stellenbosch, Parowvallei, South Africa*
John H. Andreae, *Department of Electrical Engineering, University of Canterbury, Christchurch 1, New Zealand*
Alex M. Andrew, *Department of Applied Physical Sciences, University of Reading, Reading, U.K.*
Michael A. Arbib, *Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts*
Cesar Enrique Armoza, *Faculty of Applied Mathematics, Catholic University of La Plata, Buenos Aires, Argentina*
E. Stewart Bainbridge, *Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada*
Peter Becker, *Electronics Laboratory, Technical University of Denmark, Lyngby, Denmark*
Hans Blomberg, *Systems Theory Laboratory, Helsinki University of Technology, Otaniemi, Finland.*
Kenneth E. Boulding, *Institute of Behavioral Science, University of Colorado, Boulder, Colorado*
Stein Braten, *Institute of Sociology, University of Oslo, Oslo, Norway*
I. V. Blauberg, *Institute for Systems Studies, Moscow, U.S.S.R.*
Mario Bunge, *Foundations and Philosophy of Science Unit, McGill University, Montreal, Quebec, Canada*
Peter B. Checkland, *Department of Systems Engineering, University of Lancaster, Bailrigg, Lancaster, U.K.*
C. West Churchman, *Department of Business Administration, University of California, Berkeley, California*
Roger C. Conant, *Department of Information Engineering, University of Illinois at Chicago Circle, Chicago, Illinois*
Joseph V. Cornacchio, *School of Advanced Technology, State University of New York, Binghamton, New York*
Pierre Delattre, *Saclay Center for Nuclear Studies, Gif-sur-Yvette, France*
Fred E. Emery, *Department of Sociology, Australian National University, Canberra, Australia*
Alejandro B. Engel, *Institute of Mathematics, Statistics, and Computer Science, State University of Campinas, Campinas, Brazil*
Lorenzo F. Figueras, *Department of Mathematics, University of Valencia, Valencia, Spain*
Brian R. Gaines, *Department of E. E. Sciences, University of Essex, Colchester, U.K.*
Wojciech Gasparski, *Polish Academy of Sciences, Warsaw, Poland*
Joseph A. Goguen, *Computer Science Department, University of California, Los Angeles, California*
Peter Gomez, *St. Gall Graduate School for Economics, Business and Public Administration, St. Gallen, Switzerland*
Preston C. Hammer, *Department of Mathematics, The Grand Valley State Colleges, Allendale, Michigan*
Robert E. Kalaba, *Biomedical Engineering Program, University of Southern California, Los Angeles, California*
István S. Kiss, *Hungarian Academy of Sciences, Budapest, Hungary*
Shumpei Kumon, *University of Tokyo, Tokyo, Japan*
George E. Lasker, *School of Computer Science, University of Windsor, Windsor, Ontario, Canada*
George G. Lendaris, *Systems Science Institute, Portland State University, Portland. Oregon*
Aristid Lindenmayer, *Theoretical Biology Group, University of Utrecht, Utrecht, The Netherlands*
Lars Lofgren, *Electrical Engineering Department, University of Lund, Lund, Sweden*
Mihajlo Mesarovic, *Systems Research Center, Case Western Reserve University, Cleveland, Ohio*
G. Arthur Mihram, *Department of Systems Engineering, University of Pennsylvania, Philadelphia, Pennsylvania*
Koichi Miyasawa, *Department of Economics, University of Tokyo, Tokyo, Japan*
Heiner Müller-Merbach, *Darmstadt Institute of Technology, Darmstadt, Germany*
Renato Pennacchi, *IBM Italy, S.p.A., Rome, Italy*
Franz Pichler, *Department of System Theory, University of Linz, Linz, Austria*
Brian Porter, *Department of Mechanical Engineering, University of Salford, Salford, U.K.*
Reza Razani, *School of Engineering, Pahlavi University, Shiraz, Iran*
Robert Rosen, *Dalhousie University, Halifax, Nova Scotia, Canada*
Harold A. Rothbart, *California State College, Dominquez Hills, 1000 E. Victoria Street, Dominquez Hills, California 90747*
Vadim N. Sadovsky, *Institute for Systems Studies, Moscow, U.S.S.R.*
Prem S. Satsangi, *Electrical Engineering Department, Indian Institute of Technology, New Delhi, India*
Kasumu Salawu, *Department of Business Administration, University of Lagos, Lagos, Nigeria*
Robert Trappl, *Institute for General and Comparative Physiology, University of Vienna, Vienna, Austria*
Heinz Von Foerster, *P.O. Box 399, Pescadero, California 94060*
A. Wayne Wymore, *Systems and Industrial Engineering Department, University of Arizona, Tucson, Arizona*
Lotfi A. Zadeh, *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California*
Bernard P. Zeigler, *Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel*
Hans J. Zimmermann, *Department of Operations Research, Technical University of Aachen, Aachen, Germany*
Pranas Zunde, *School of Information and Computer Science, Georgia Institute of Technology, Atlanta, Georgia*

The *International Journal of General Systems* is an interdisciplinary periodical devoted primarily to the publication of original research and educational contributions relevant to general systems. Tutorial and survey articles, reviews, bibliographies, short notes and letters, and a calendar of forthcoming activities related to general systems will be included. Basic areas covered by the journal are: Foundations of general systems theory and methodology, applications of the methodology in various branches of science, technology, humanities, and the arts, general systems philosophy, and general systems education. Examples of topics within these areas are: Principles of modeling and simulation, systems analysis and synthesis, optimization, identification, general principles of experimentation, problems associated with extremely complex systems, studies concerned with various classes of systems such as self-organizing, adaptive, self-producing, fuzzy, hierarchical, etc. In order to cover the diverse activities in general systems research and education throughout the world, an editorial board with international representation has been selected. Three major criteria determine the selection of submitted articles: originality (except for invited tutorial articles), high professional quality, and direct relevance to general systems research or education.
Books for review should be sent direct to the editor at the address above, and not to the publishers.

SUBSCRIPTION RATES Per volume of four issues, postpaid

	USA/elsewhere	*Great Britain*
Library/Institution/Company	$86.50/£48.00	£42.50
Individuals whose institutions already subscribe	$28.00/£15.50	£10.00

Subscription rates for USA/elsewhere include a distribution charge of $12.00 for postage and handling and air freight to USA and Canada

Subscriptions should be sent to Gordon and Breach Science Publishers Ltd., 42 William IV Street, London WC2, U.K.
JANUARY 1979 issue.

© 1979 Gordon and Breach Science Publishers Ltd., 42 William IV Street, London WC2. Printed in Great Britain by Bell and Bain Ltd., Scotland. Mailed in the U.S.A. by Expediters of the Printed Word Ltd., 527 Madison Avenue, New York, NY. 10022, U.S.A.

t. J. General Systems
79, Vol. 5, pp. 1–12

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

APPROXIMATE, NON-DETERMINISTIC MODELLING OF BEHAVIOUR SEQUENCES

IAN H. WITTEN

Man-Machine Systems Group, Department of Electrical Engineering Science, University of Essex, U.K.

(*Received June 6, 1977: in final form March 2, 1978*)

This paper describes a technique for forming approximate, non-deterministic, finite-memory models of behaviour sequences. The models depend on a "memory-length" parameter k which controls their complexity and hence goodness-of-fit to the given behaviour by constraining them to reproduce correctly all sub-sequences of length k and less in the behaviour. This leads to a simple, incremental, efficient, and surprisingly effective modelling method. Inputs and outputs to the system being modelled are not distinguished but can be intermingled freely in the behaviour sequence, and the extent to which the input process cannot be modelled causally will be reflected in the number of non-deterministic transitions in the model. Examples are included of the use of the technique.

INDEX TERMS System identification, behaviour-to-structure transformation, finite-memory models, non-deterministic automata, incremental modelling, grammatical inference, admissible models, length-k modelling.

INTRODUCTION

behaviour sequence of discrete events can be ored in many different ways. If the sequence is ındom, the most efficient way is not to code it all, but simply to remember the sequence self.[1] However, a sequence which is generated y a deterministic finite-state automaton has a ructure which can easily be exploited to dis-ɔver an effective code.[2] Indeed, if the automaton input-less, its output must eventually be per-ɔdic, and of course one copy only of the periodic ehaviour need be stored. Most behaviour se-uences encountered in practice lie between these vo extremes.

A behaviour sequence can be summarized in a ay which is often useful for predicting the next w events in it but which does not retain com-ete information about it, by constructing a non-eterministic model.[3] Consider the sequence BCDABBCDBCDABC. A minimal-state Moore odel with non-deterministic transitions is shown Figure 1a. The price paid for such a simple odel is its ability to generate sequences which e not part of the particular one used to con-ruct it—the sequence EBBBBC, for example. ut notice that the model can be built very mply and added to incrementally as the event quence continues to unfold. We call this model, which has a state for each different symbol occurring in the behaviour and a link between two states only when the output of the first is followed, in the sequence, by that of the second,

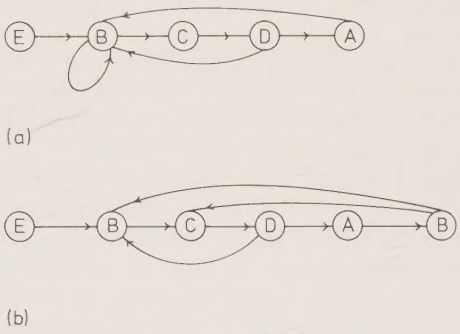


FIGURE 1 Non-deterministic models of the sequence EBCDABBCDBCDABC. (a) Ingenuous model (length-2). (b) Length-4 model.

the *ingenuous* non-deterministic model of the sequence. For a behaviour with less structure, such as EADCEEDDBCDEDBA, the ingenuous model is more heavily connected (Figure 2a), and as the sequence unfolds further one would expect more connections to be added, rendering it less useful for prediction.

This paper explores the effect of strengthening the structure of such models by adding the constraint that they must be incapable of generating sequences of length k or less that do not occur in the observed behaviour, for some specified integer k—the *length-k constraint*. This increases the state count of the model and decreases its non-determinism. Figures 1b and 2b show $k=4$ models for the well- and ill-structured

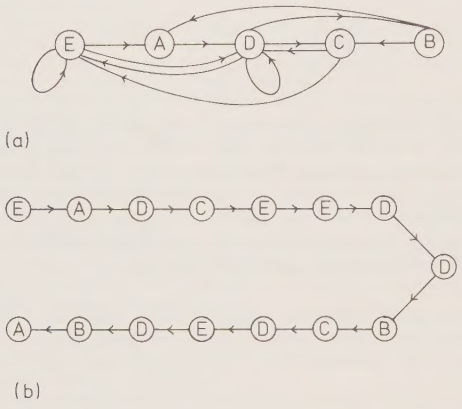


FIGURE 2 Non-deterministic models of the sequence EADCEEDDBCDEDBA. (a) Ingenuous model (length-2). (b) Length-4 model.

sequences above. If the behaviour is generated by an n-state deterministic automaton, and k is chosen to be at least as great as n, then a length-k model of the sequence will be exact and deterministic. Note that the ingenuous model of a behaviour sequence is a length-2 model.

Length-k models are *finite-memory* models,[4] because they only store information about k-tuples of the behaviour. In the deterministic case, length-k finite-memory models can be reduced to equivalent minimal finite-state ones by first constructing a non-minimal model whose states are simply the k-tuples which occur in the behaviour sequence, with the appropriate transitions, and then minimizing it by the usual technique of Nerode equivalence.[2] However, no simple reduction algorithm has been found to yield minimal-state non-deterministic automata from non-minimal ones.[5]

There are, of course, other ways of forming non-deterministic models of behaviour sequences. Given a model space, Gaines[3] introduces an admissible set of models, which depends on a measure of poorness-of-fit of a model to a behaviour sequence, and a criterion of model simplicity. A model is *admissible* if it cannot be improved in simplicity without increasing poorness-of-fit, and it cannot be improved in goodness-of-fit without increasing complexity. In general, there will be a collection of admissible models of varying simplicity and poorness-of-fit. This admissible set of models can be found by enumerating the elements of the model space (assuming this is possible), and testing each model for goodness-of-fit to the observed behaviour sequence. This procedure is, of course, very slow.

The length-k criterion itself provides a measure of how well a model fits a behaviour sequence. Defining an *extraneous sequence* of events as one which can be produced by the model but does not occur as a subsequence of the observed behaviour, then the goodness-of-fit of the model to the behaviour is the length of the shortest extraneous sequence that it can generate—$k+1$, in the case of a length-k model. For a measure of model complexity, take, as usual, its state count. Then the set of minimal-state length-k models, for $k=2,3,\ldots$, is exactly the admissible subspace.

Call the subsequences within a sequence the *words* of the sequence. Then, following Lempel and Ziv,[6] define the *vocabulary* of a sequence to be the set of all its words. The *length-k vocabulary* is the set of all words whose length is k. It is a fact that the length-k vocabulary of a sequence contains exactly the information needed to construct a length-k model of it. Clearly the length-k vocabulary can be reconstructed from a length-k model by tracing all length-k sequences from each state. No extraneous k-sequences can be produced, and we must be able to trace the entire behaviour sequence through the model, so all of its length-k words must be generable by this procedure. It is shown in section 5 how a length-k model can be constructed from the length-k vocabulary. (Of course, *some* length-k models contain more information about the sequence—after all, the model formed by storing the behaviour itself as a succession of states is length-k). Thus, the length-k procedure is tantamount to modelling a behaviour as a *strictly k-testable event*.[7]

There is a close correspondence between the problem of modelling behaviour sequences as automata and the inference of finite-state grammars from examples.[8,9] For the latter, one is usually presented with several examples of sequences, and has to infer a grammar which

nerates them. This can be accomplished within framework described here by introducing a cial delimiter character into the behaviour quence, where the symbols before the delimiter known to be of no assistance in predicting se which follow it. Then, several examples of naviour can be separated by delimiters and orporated into one sequence.

The behaviours which are considered here are nply sequences of events, or symbols. No disction is made between inputs and outputs. If sequence is produced by an input–output achine, and contains inputs and outputs alterely, then the length-k modelling procedure l attempt to model the process that is produc- the inputs as well as the behaviour of the achine itself. If the inputs are chosen randomly d the behaviour sequence is long enough—in sense which can be made precise and depends the parameter k of the modeller—for this to apparent), the input process will simply be delled non-deterministically. In cases where k sufficiently large for the state-structure of the achine to be identifiable from length-k windows the input–output sequence, each of its states l be faithfully represented in the length-k mo-, and will have non-deterministic branches to ates" representing each possible input symbol. t is known that the machine is deterministic, it a simple matter to re-configure such a nonterministic output-only model of it into a deministic input–output model. This can be done tomatically, without knowledge of which symls are inputs and which are outputs—indeed, input and output alphabets need not be joint.

Biermann and Feldman[10] have shown how the rode equivalence relation can be expanded to oduce a spectrum of models which approhate an input–output function, defined on strgs of symbols, from example input–output rs. Their work is similar to that described here that it yields length-k models. However, by nsidering a single sequence of events, instead input–output pairs, we achieve a simpler and re uniform presentation in which inputs and tputs need not be distinguished but are simply erleaved in the behaviour sequence.

The next section defines length-k models prely, and introduces some related concepts. To preciate the difficulty of finding minimal-state gth-k models, it is important to be familiar h some negative properties of such models which are presented in the subsequent section. Section 4 discusses a result which can often be used to reduce non-minimal-state models. Unfortunately, because of the difficulties hinted at above, it cannot guarantee to produce a minimal-state version. A canonical form of length-k model is introduced in Section 5, together with a simple procedure for forming it. While it is not necessarily a minimal-state model, it is usually close to minimal-state. In addition, it has a nice property of observability that is not shared by the minimal-state model. It is in keeping with most adaptive strategies to identify behaviour incrementally, instead of waiting until an entire identification experiment has been performed.[11] This can be done by determining at each stage the admissible models which account for the behaviour observed so far. Fortunately, the length-k model admits of incremental modification when a new event is appended to the behaviour sequence which does not fit the current model. The incremental procedure for updating the model is developed in Section 6. A simple, fast, but rather space-inefficient method of implementing length-k models in the form of hash tables is introduced in the next section. Finally, some examples of length-k models are discussed.

2. LENGTH-K MODELS

The models we will consider take the form of non-deterministic Moore automata, with state set S, output set O, a set-valued next-state function $\mu: S \to 2^S$, and an output function $\lambda: S \to O$. O has one distinguished element, which will be denoted by "/", called the *delimiter*. $\mu(s)$ is called the *successor set* of state s. We can define μ's inverse $\mu^{-1}: S \to 2^S$ by

$$t \in \mu^{-1}(s) \quad \text{if and only if} \quad s \in \mu(t).$$

$\mu^{-1}(s)$ is the *antecedent set* of s.

If the current state in the model is s, then the next state may be any member of its successor set $\mu(s)$. Thus in state s, the model predicts that the next output will be $\lambda(t)$ for some $t \in \mu(s)$. The size of $\mu(s)$ governs the extent to which the automaton is a successful model of the behaviour sequence: in the extreme case, if $\mu(s)$ has just one element for each s, then it is a deterministic model. However, a premise of this work is that it is very often useful to accept non-determinism in

models, partly to limit the complexity of the model (for example, when learning about a real-world situation[12]), partly to avoid the disastrous effects of noise on deterministic modelling procedures,[13] partly because the machine which generated the behaviour may itself be non-deterministic (as is the case in grammatical inference[8,9]), and partly because it allows one to cope easily and naturally with unpredictable inputs to the model.

A *behaviour* is a finite sequence of symbols drawn from O, the automaton's output set. An automaton is said to *model* a behaviour $B = o_1 o_2, \ldots, o_n$ if the sequence can be traced through the model; that is, if there exists a sequence of states $s_1, s_2, \ldots, s_n \in S$ with

$$s_{i+1} \in \mu(s_1) \quad 1 \leqq i \leqq n-1 \quad \text{and} \quad \lambda(s_i) = o_i \quad 1 \leqq i \leqq n.$$

Note that some backtracking may be necessary in tracing a behaviour through the model, as when, for example, a state is reached which has two states in its successor set both with the same output. It seems desirable to favour models which are *followable* in the sense that

$$u, v \in \mu(s) \Rightarrow \lambda(u) \neq \lambda(v).$$

A behaviour sequence can always be traced through a followable model without backtracking.

An automaton is a length-k model of a behaviour B if it is a model of it, and, further, every sequence of k outputs which it can generate occurs somewhere in B. Of special interest are minimal-state length-k models, or, at least, models which are close to minimal state.

A final concept which will be useful is that of *m-observability*. A model is m-observable if any sequence of m consecutive symbols from a behaviour which it models uniquely determines the state of the model which generates the last symbol of the sequence. Thus, if two sequences of states in the model can produce the same sequence of m outputs, the last states in the sequences must be the same. If a model is not followable, then it cannot be m-observable for any m.

To illustrate these definitions, Figure 3 shows four examples of length-4 models of the behaviour SFCRGFFCRFCRGFCR. None of the first three are length-5 or greater, since they can all generate the extraneous FFCRG. Of course, a length-k model is necessarily a length-$(k-1)$ one, so they are all length-2 and length-3 models. T second and third are minimal-state. The third not followable, for the G state has two success F states, but the others are. Only the first two 3-observable; moreover, the second is observable.

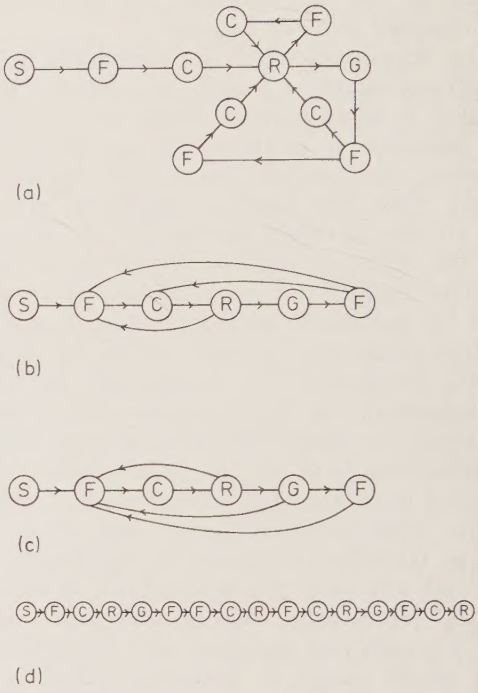


FIGURE 3 Length-4 models for the sequ SFCRGFFCRFCRGFCR.

3. NEGATIVE RESULTS

Minimal-state length-k models are difficult discover without resorting to lengthy search. appreciate why this is so, one should be famil with some negative properties of such mod which, although they can be demonstrated simple counter-examples, do not become appar when modelling the majority of behavio sequences.

A. Consider the problem of forming a lengt model of a behaviour B *incrementally*, as symbols of the behaviour appear one by o Denote by M_r the minimal-state lengt model of the first r symbols of B. Then number of states in M_r may decrease as r creases. In other words, the model may beco simpler as more of the behaviour is seen.

Notice that this cannot happen in determini modelling. Viewed deterministically, the co

kity of a sequence is a non-decreasing function ts length. However, with length-k modelling, re observations may reveal that the be-iour is too complex to be handled determin-cally by a length-k model, and this realization the model must become non-deterministic lead to a considerable collapse of states. nsider, for example, the sequence AABBBABBAAA. For the initial part of it, a gth-3 model needs three states. Only as the uence unfolds further does it become apparent the two-state ingenuous model is a length-3 . Or suppose we attempt, say, length-12 in-mental modelling of a random sequence of A's, C's, D's, and E's. The model can be expected grow to some considerable size before it nks to the inevitable ingenuous one.

his should not be interpreted as a disadvan-e of length-k models. It is exactly these cir-nstances that cause a deterministic modelling cedure to blow up.[13] At least length-k models w when to give in!

. There does not necessarily exist a minimal-e followable length-k model for a behaviour uence.

s mentioned in the previous section, it seems rable to use a followable model whenever sible, to avoid the necessity to backtrack en tracing a behaviour through the model. hniques for incrementally modifying models he light of new behaviour naturally tend to d followable models: why create a new state en one is there already which will do? thermore, the minimal-state model in Figures 2b, and 3b are all followable. However, ure 4a shows the minimal-state model of the uence XABCDBCDEABCFBCF, and it is not owable. Figure 4b shows the corresponding imal-state followable model, which has one e state.

. The minimal-state followable length-k mo-of a behaviour is not necessarily $(k-1)$-ervable.

s shown below, there are simple techniques forming $(k-1)$-*observable length-k* models of aviours. In most cases, these form minimal-e length-k models—for example, with the be-iours of Figures 1, 2, and 3. However, neither he models of Figure 4a and 4b are 3-observ-. The minimal-state 3-observable length-4 del for this sequence is given in Figure 4c, and two more states than the minimal model.

his result presents further problems for in-cremental modelling methods which attempt to retain only minimal-state models. When trying to identify the target state for a new transition that has proved necessary, one naturally looks ahead

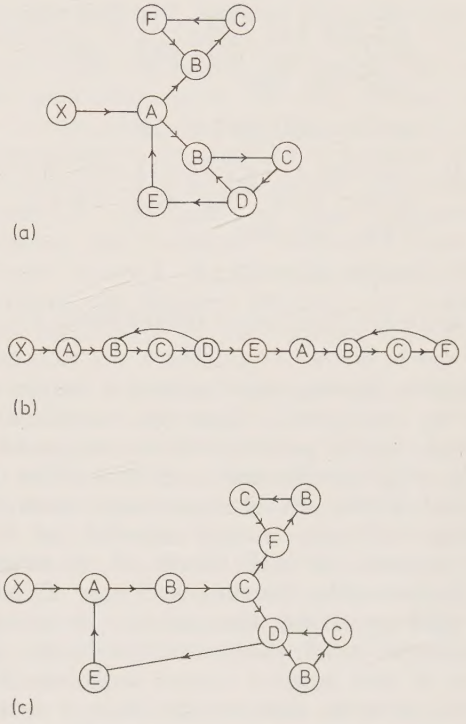


FIGURE 4 Length-4 models for the sequence XABCDBCDEABCFBCF. (a) Minimal-state, non-followable (9 states). (b) Minimal-state, followable (10 states). (c) Minimal-state, 3-observable (11 states).

at the next $k-1$ elements of the behaviour sequence and searches for a model state which replicates these. Result C shows that such a state will not necessarily by unique.

4. REDUCTION OF NON-MINIMAL-STATE MODELS

It seems to be difficult to produce a neat condition which characterizes minimal-state models in a way which allows one to reduce non-minimal-state models into minimal-state equivalents. The best reduction theorem obtained as yet is:

THEOREM *A necessary condition that a length-k model is minimal-state is: for any* $s_1, s_2, \ldots, s_{k-1}$; $s'_1, s'_2, \ldots, s'_{k-1} \in S$ *with*

A. $s_{i+1} \in \mu(s_i)$ $1 \leqq i < k-1$;

B. $s'_{i+1} \in \mu(s'_i)$ $1 \leqq i < k-1$;

C. *for any* i $(1 < i \leqq k-1)$ *such that there exists* $s \in \mu^{-1}(s_i)$ *with* $s \neq s_{i-1}$ *and* $s \notin \mu^{-1}(s'_i)$, *then*

$$\mu(s_j) \supseteq \mu(s'_j) \quad i \leqq j \leqq k-1;$$

and *D.* *for any* i $(1 < i \leqq k-1)$ *such that there exists* $s' \in \mu^{-1}(s'_i)$ *with* $s' \neq s'_{i-1}$ *and* $s' \notin \mu^{-1}(s_i)$, *then*

$$\mu(s'_j) \supseteq \mu(s_j) \quad i \leqq j \leqq k-1;$$

then $\lambda(s_i) = \lambda(s'_i)$ $1 \leqq i \leqq k-1$ implies that $s_i = s'_i$ $1 \leqq i \leqq k-1$.

Proof by showing that if state s_i is merged with state s'_i, $1 \leqq i \leqq k-1$, then no additional k-sequences can be generated by the new model.

The rather complicated conditions of the theorem test if two sets of states exist which both produce the same output sequence, of length $k-1$ symbols—in other words, if the model is $(k-1)$-observable. Conditions C and D ensure that there are no side branches into or out of the sequences of states which would permit the generation of new length-k output sequences if the two sets of states were merged. That it does not constitute a *sufficient* condition for the model to be minimal-state can be seen by examining the two length-4 models of Figure 5. Model 5b, while

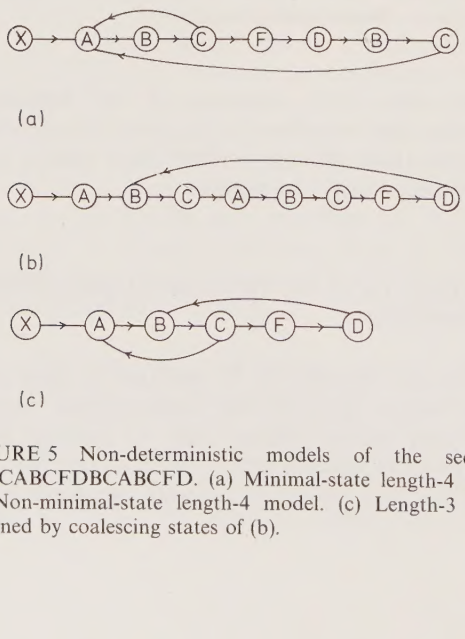


FIGURE 5 Non-deterministic models of the sequence XABCABCFDBCABCFD. (a) Minimal-state length-4 model. (b) Non-minimal-state length-4 model. (c) Length-3 model obtained by coalescing states of (b).

clearly not minimal-state, cannot be reduced b the theorem. If we identify the states produci the two ABC-sequences with s_1, s_2, s_3; s'_1, s'_2, s then condition C is not satisfied, for there exis $s \in \mu^{-1}(s_2)$ with $s \neq s_1$, $s \notin \mu^{-1}(s'_2)$, b $\mu(s_3) \not\supseteq \mu(s'_3)$.

It is not the case that the conditions are simp too strong, for here states s_i and s'_i cannot merged without destroying the length-4 proper of the model. Figure 5c shows the result merging, and this model can produce the e traneous 4-sequence DBCF.

A very useful corollary is obtained from t theorem by taking s_i and s'_i to be the same sta for $1 \leqq i \leqq k-2$ and for $2 \leqq i \leqq k-1$.

COROLLARY *Given a minimal-state length-k m del, then for any* $t \neq \mu \in S$,

A. $\mu^{-1}(t) = \mu^{-1}(u)$ *implies* $\lambda(t) \neq \lambda(u)$;

B. $\mu(t) = \mu(u)$ implies $\lambda(t) \neq \lambda(u)$.

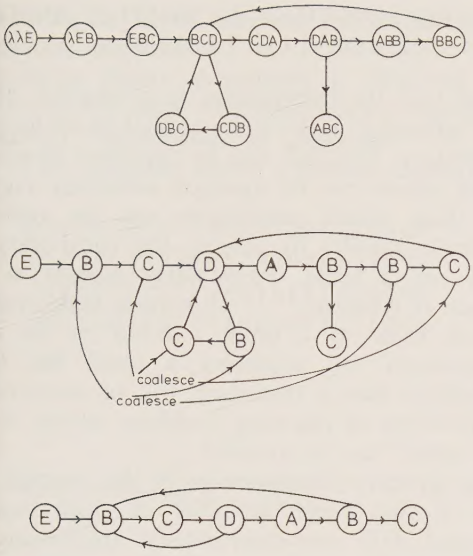
In other words, a minimal-state length-k mod cannot have two distinct states with the sa output symbol which

A. *possess the same antecedent set,*

or B. *possess the same successor set.*

5. THE CANONICAL LENGTH-K MODEL

An easy way of constructing a length-k model a behaviour is to split it into overlapping $(k-1$ tuples and form the ingenuous model of the ne sequence, by taking each of these as a state a marking transitions between states which occ consecutively. The output from a state is taken the last element of its $(k-1)$-tuple. To preser the first few elements of the behaviour, it preceded by $k-2$ null symbols (ε) before b ginning the modelling procedure.[14] Delimiters a handled by mapping any symbols which prece a delimiter in a $(k-1)$-tuple into ε. The model obtained can be reduced, without sacrificing length-k property, by coalescing states whi have the same output and the same successor s according to part B of the corollary above. W call the result the *canonical length-k model* of t sequence. Figure 6 shows the formation of t canonical length-4 model for the sequence Figure 1.



GURE 6 Forming the canonical length-4 model for the uence EBCDABBCDBCDABC.

HEOREM *The canonical length-k model of a quence is* $(k-1)$*-observable, and hence llowable.*

Proof (informal) If it were not $(k-1)$-servable, there would exist two sets of states s_1, $\ldots, s_{k-1}$ and $s'_1, s'_2, \ldots, s'_{k-1}$, with transitions $\in \mu(s_{i-1})$ and $s'_i \in \mu(s'_{i-1})$, which both produce e same sequence of outputs, but end in different tes (so $s_{k-1} \neq s'_{k-1}$). This cannot happen in the genuous model of $(k-1)$-tuples, for in this each te is characterized by the $k-1$ outputs which led it, and there is a unique state for each $(k-1)$-ple which occurs in the behaviour. Neither n the reduction process cause it to happen, for ly states with the same output are coalesced. ote that $(k-1)$-observability implies llowability.

It is apparent from these properties of the nonical model, and results B and C of ction 3, that it is not necessarily a minimal-te length-k model. Also, it cannot be further duced by the theorem of Section 4, since it is $-1)$-observable and so part B of the corollary he only case of the theorem that can apply to it, d this has already been used to reduce it as ch as possible. It appears that in practice, wever, the model is very often minimal-state, d always close to it. Unfortunately, because of the apparently intractable nature of the minimal-state model, we have not yet succeeded in obtaining a reasonably tight upper bound to the amount by which the state count of the canonical model can exceed that of the minimal model.

An alternative model to the canonical one can be constructed by splitting the behaviour into overlapping $(k-1)$-tuples and forming the ingenuous model as before, but taking the output from each state to be the *first* element of its $(k-1)$-tuple, instead of the last, $k-2$ nulls should be appended to the sequence to preserve the last few elements of it in the model. The model so formed is the same as that proposed by Biermann and Feldman.[10] (However, their k is the same as our $k-2$). It can be reduced, without sacrificing the length-k property, by coalescing states which have the same output and the same antecedent set, according to part A of the corollary above. Indeed, a whole suite of models can be obtained by taking the output from each state of the ingenuous model of $(k-1)$-tuples to be the i'th element of its tuple, for $1 \leqq i \leqq k-1$. The behaviour sequence should be preceded by $i-1$ nulls and followed by $k-i-1$ nulls to preserve its ends in the model. However, the canonical model is preferred because it is the only one which is $(k-1)$-observable and hence followable. This small advantage is not offset by any disadvantages.

Note that the construction procedure for the canonical length-k model does not need the full behaviour sequence to be given—merely its length-k vocabulary is sufficient. This contains full information about transitions between $(k-1)$-tuples of the behaviour sequence.

6. INCREMENTALLY UPDATING THE CANONICAL LENGTH-K MODEL

The canonical length-k model for a behaviour can be updated easily if more of the sequence appears which does not fit the model. One need only remember the last k symbols of the new behaviour. This gives length-k modelling a great advantage over other methods of non-deterministic modelling which require the complete sequence to be stored, and re-model it from scratch when the current model proves inadequate.

Suppose first that the ingenuous model of $(k-1)$-tuples, before reduction, for the behaviour $\ldots o_2 o_3 \ldots o_{k-1}$ is given, but that the next event, o_k, does not fit the model. If a state characterized by $o_2 o_3 \ldots o_k$ does not exist already in the model, then one must be created. Note that the length-$(k-1)$ output sequence which characterizes a state can be found by tracing backwards from that state along transitions and concatenating in reverse order the outputs of the states so traced. If a delimiter is encountered, either in the behaviour or when tracing back through the model states, the elements which precede it in its $(k-1)$-tuple are mapped to ε. Then, a transition link must be made from the state characterized by $o_1 \ldots o_{k-1}$ to that characterized by $o_2 \ldots o_k$. This simple procedure is sufficient to construct the ingenuous model of $(k-1)$-tuples for the augmented behaviour.

Now consider how to update the canonical length-k model, which is a reduced version of the ingenuous model of $(k-1)$-tuples, in the light of the new event o_k. Again, if a state characterized by $o_2 o_3 \ldots o_k$ does not exist already in the model, one must be created. Because the model is $(k-1)$-observable, if one does exist it is unique. A difficulty arises because the state characterized by $o_1 o_2 \ldots o_{k-1}$ may also be characterized by a different sequence, say $o'_1 o'_2 \ldots o'_{i-1} o_i \ldots o_{k-1}$, because of the reduction process that the model has undergone. If this were so and we simply created a link from this state to the one characterized by $o_2 \ldots o_k$, the model would be capable of generating extraneous k-sequences of the form $\ldots o'_{i-1} o_i \ldots o_{k-1} o_k$. The difficulty is resolved by expanding part of the model by creating duplicate copies of the relevant states which were coalesced in the reduction process; adding the new transition, and reducing again.

Formally, suppose the sequence $o_1 o_2 \ldots o_{k-1}$ is generated by states $s_1 s_2 \ldots s_{k-1}$. Let i be the smallest index with $\mu^{-1}(s_i) \neq \{s_{i-1}\}$ and $\lambda(s_i) \neq$ "/", the delimiter. Remove s_i from $\mu(s_{i-1})$. Create new states $s'_i, s'_{i+1}, \ldots, s'_{k-1}$ with $s'_i \in \mu(s_{i-1})$, $\mu(s'_i) = \{s'_{i+1}\}$, $\mu(s'_{i+1}) = \{s'_{i+2}\}, \ldots, \mu(s'_{k-2}) = \{s'_{k-1}\}$, and create a link from s'_{k-1} to the new state characterized by $o_2 o_3 \ldots o_k$. Then reduce the new model according to corollary B of Section 4, if possible.

It is not difficult to see that this procedure augments the canonical length-k model of the behaviour $\ldots o_1 o_2 \ldots o_{k-1}$ to that of the behaviour $\ldots o_1 o_2 \ldots o_{k-1} o_k$, as required.

7. IMPLEMENTING LENGTH-K MODEL IN THE FORM OF TABLES

Recall that the information in a length-k mod of a behaviour is the same as that in its length- vocabulary. This fact can be exploited to yield model which can be updated extremely rapidl and from which predictions can be retrieve extremely rapidly by storing the vocabulary d rectly, as a table. Particularly elegant is th method of Andreae,[12,15] who stores each event i a hash table at a point indexed by its $k-$ predecessors. By accepting a small but finit probability that a transition will be overwritte the problem of chaining overflow entries in th hash table[16] can be avoided.

The primary disadvantage of this storage me thod is that large amounts of store can b occupied with non-deterministic transitions. I the extreme case, for length-k modelling of random sequence whose elements are drawn fro an alphabet of α symbols, α^k events must b stored in the hash table, and so it must hav considerably more than this number of cell There is an irony here in that large areas of sto are used to hold events which will be of little u for prediction anyway—simply because they a nondeterministic.

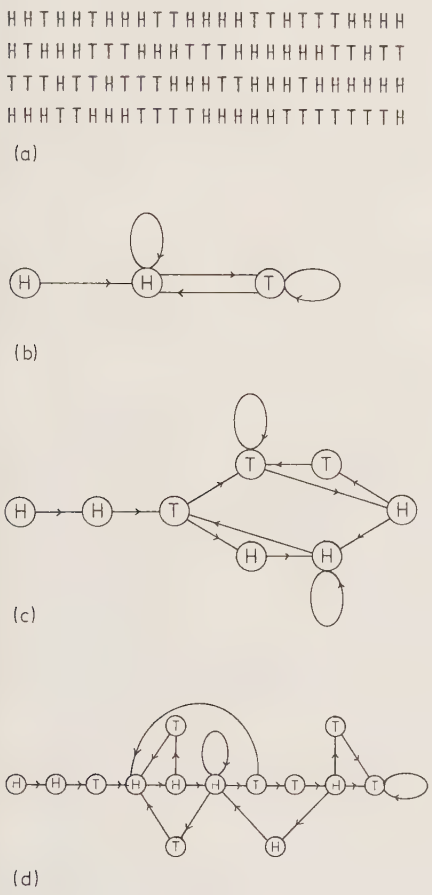
However, storage in a hash table avoids th searching which must be done to find the ne state, whenever a canonical length-k model updated. Retrieval and updating times are ind pendent of the store size. Perhaps most impo tantly, for large models, the uniform storag structure of a hash table obviates the need fo time-consuming free-space and pointer operation on secondary storage media which characterize the maintenance of large list-structured model The hashing mechanism has the further interestin property that, because of its pseudo-random na ture, the memory of the model is distribute throughout the store—if a part is erased, a fairl uniform degradation of predictive power take place throughout the model. This parallels th "holographic" property of the human brain' storage mechanism.

8. EXAMPLES OF LENGTH-K MODELLING

Gaines[3] applies his exhaustive-search non deterministic modelling procedure ATOM to s

al examples. Here, length-k models for some these are examined.

Figure 7a shows a sample of observed behaviour of 100 heads and tails from a stochastic tomaton introduced by Andreae[12] in the m of an anecdote about a gaming club owner



URE 7 (a) Behaviour sequence from a coin-tossing hine. (b) Length-2 and length-3 canonical model. (c) Length-4 onical model. (d) Length-5 canonical model.

o suspects a coin-tossing machine of being ged. The canonical length-2, length-3, length-4, l length-5 models are shown. The bias of the chine is not apparent in the length-2 and gth-3 structure of the behaviour sequence. The gth-4 model has two states with deterministic nsitions, showing that some symbols are perly predictable. Stripped of the two leading-in tes,[17] it is equivalent to Gaines's admissible 4-te Mealy model.

If the sequence were actually generated by this machine, and the observed behaviour was long enough, one would expect the length-5, length-6, ..., models to be the same as this. In fact, the length-5 model differs from the length-4 one because the sequences TTHHT and THHTT do not occur in the behaviour, although they can be generated by the length-4 model. However, a longer behaviour would either reveal that this omission is fortuitous, or indicate that the machine actually had the more complicated structure of Figure 7d.

It is interesting to estimate just how long one might expect to wait before all k-tuples from a random source occur. Suppose the alphabet contains α symbols, then there are α^k different k-tuples. Treating the occurrence of each as an independent event, suppose n of the α^k have been encountered. Denote the expected time until the next is seen by m. Then

$$m=\frac{\alpha^k-n}{\alpha^k}+\frac{n}{\alpha^k}(1+m),$$

and so $m=\dfrac{1}{1-n\alpha^{-k}}$.

Thus the mean time for all the k-tuples to be seen is

$$1+\frac{1}{1-\alpha^{-k}}+\frac{1}{1-2\alpha^{-k}}+\cdots+\frac{1}{1-(\alpha^k-1)\alpha^{-k}}$$
$$=\alpha^k\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{\alpha^k}\right)$$
$$\doteq\alpha^k(k\log_e\alpha+0.577).$$

How often do k-tuples appear? Certainly, a fresh tuple occurs with each symbol of the behaviour: however, successive tuples will be correlated because they overlap. A new, independent, k-tuple appears every k symbols. Thus the expression derived is a lower bound to the mean length of behaviour sequence before every k-tuple has appeared, and an upper bound is obtained by multiplying it by k.

This reasoning does not apply to the behaviour sequences generated by the length-4 model of Figure 7c, for this is not a pure random source, and, in particular, we know that some k-sequences cannot occur. It could no doubt be

modified to fit the model properly. However, for a rough estimate of the mean length of sequence before all possible 5-tuples occur, these considerations can be ignored, and substituting $\alpha=2$, $k=5$ in the above expression leads to a lower bound of 130 and an upper bound of 650. Thus it is not surprising that the length-100 behaviour sequence of Figure 7a does not contain all 5-tuples that the length-4 model can generate.

A grammatical inference example which has been used by several authors[3,8,9] is given in Figure 8. Seven sample strings of the language are shown, separated by delimiters, together with length-2, length-3, and length-4 models for the sequence. The first model corresponds to the grammar

$$S\rightarrow CA\alpha|CB\beta|B\beta$$
$$\alpha\rightarrow A\alpha|B\beta$$
$$\beta\rightarrow B\beta|A\alpha|\lambda,$$

and the second to

$$S\rightarrow C\alpha|BB\beta$$
$$\alpha\rightarrow A\alpha|B$$
$$\beta\rightarrow A\alpha|B\beta|\lambda$$

(λ is the empty string.)

The third removes the recursion, and is the minimal-state non-deterministic model which generates exactly the strings given:

$$S\rightarrow CAAAB|BBAAB|CAAB|BBAB|CAB|BBB|CB.$$

None of the models produces Evans's[8] grammar

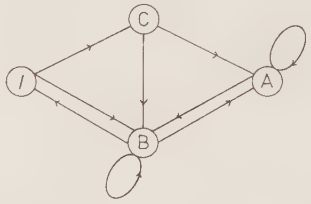
$$S\rightarrow C\alpha|BB\alpha$$
$$\alpha\rightarrow A\alpha|B$$

which includes A-recursion but no B-recursion: however, examination of the strings seems to indicate that there is as much evidence for B-recursion as there is for A-recursion! The discrepancy between Evan's grammar and the length-k models is probably attributable to the small amount of data on which both are based.

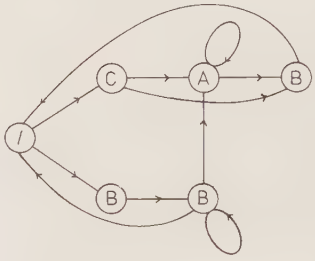
Gaines discusses the inference of a sorting program from a trace of the process as applied to one example. We will not reproduce details of the trace here. Suffice it to say that a length-4 mod corresponds exactly with a flowchart of the pr cedure, non-deterministic branches being place correctly at decision points.

/CAAAB/BBAAB/CAAB/BBAB/CAB/BBB/CB/

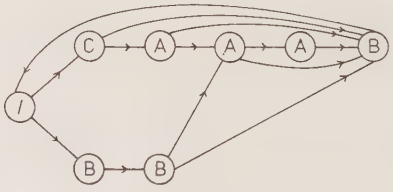
(a)



(b)



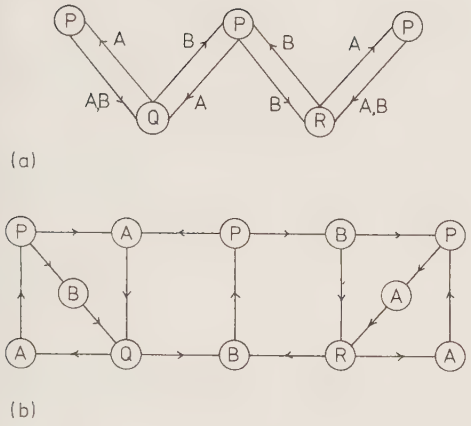
(c)



(b)

FIGURE 8 (a) Seven samples of a language, separated b delimiters (/). (b) Canonical length-2 model. (c) Canonic length-3 model. (d) Canonical length-4 model.

As noted in the Introduction, a behaviou sequence containing inputs can be modelled b simply treating the inputs as elements of th sequence. If the environment is deterministic, length-k model usually reveals which symbols ar inputs and which are outputs because the forme are preceded by non-deterministic branches. Th canonical length-5 model of a long input-outpu sequence generated from the machine Figure 9a is shown in Figure 9b. Five is th minimum length of input–output sequence whic can capture the difference between the three "F states.



GURE 9 (a) Environment with inputs (A, B) and outputs Q, R). (b) Canonical length-5 model of a long sequence of form PAQAPAQBPBRAPBRAPARBPAQA....

CONCLUSIONS

ıe transitions from considering infinite samples continuous behaviour to finite samples of screte behaviour, and from linear systems to n-linear systems, are interesting features of odern work in engineering and related disciples. Recently, there has been evidence of a trend wards simplifying the problems posed by nonearity in finite systems by considering only the cal contexts of events in a behaviour sequence stead of trying to deal with the whole beviour at once—in areas as diverse as automata eory,[7] language identification,[10] complexity eory,[6] and learning machine practice.[12,15] The ork described here treats the problem of haviour-to-structure transformation in this new ht by considering only the effects of interacns between events which are less than a fixed stance apart in the behaviour sequence. This ads to a simple, incremental, efficient, and rprisingly effective modelling technique-length-modelling.

The methodology of Gaines[3] for identification th non-deterministic models has been adopted. is introduces the concept of an admissible set of odels as the solution to the identification proem, rather than a single model, to accomodate the inevitable trade-off between model nplicity and goodness-of-fit. The length-k criteon provides a convenient measure of the latter, ile for the former, the model's state count is ed. However, examination of certain properties of minimal-state length-k models—the negative results of Section 3—indicates that they are rather difficult to find without excessive searching, and especially difficult to model incrementally. We avoid falling into the "optimality trap" of excessive and over-zealous concern with optimality at the expense of getting the job done, exemplified by much contemporary work on identification and control, by defining a convenient, unique, easy-to-form "canonical length-k model" for a behaviour sequence, which is in fact often the minimal-state model and always seems to be close to it in practice. However, a reasonably tight upper bound for the amount by which the complexity of the canonical model can exceed that of the minimal one has yet to be discovered.

The modelling procedure takes as input a single behaviour sequence of events, or symbols. A distinguished "delimiter" symbol permits the concatenation of several independent samples of behaviour into one uniform sequence, and is handled appropriately by the procedure. Inputs and outputs to the system being modelled can be intermingled freely in the behaviour sequence; the extent to which the input process cannot be modelled causally will be reflected in the number of non-deterministic transitions in the model. It is this uniformity which distinguishes the current work from the similar, fixed-context models of Biermann and Feldman.[10] It should be noted that structure will always be found in short sequences, and so a portion of behaviour which is sensibly long in comparison with k, the model parameter, is needed for non-determinism to be discovered. The considerations of Section 8 lead to a figure of between $k.\alpha^k.\log_e\alpha$ and $k^2.\alpha^k.\log_e\alpha$ for the length of a purely random behaviour sequence which is to be modelled correctly, where α is the cardinality of the alphabet of symbols. The strong dependence of the estimate upon k highlights a further dimension of the model simplicity/goodness-of-fit trade-off—the more complex the model, the longer it takes to form.

Further refinements to the modelling procedure, such as computing observed frequencies for the transitions and interpreting them probabilistically, and finding the "surprise" occasioned by each symbol of the behaviour sequence, can easily be added.[3] One topic of immediate concern is the comparison of canonical length-k models with the set of optimal causal

Moore models obtained by searching through the model space, for various examples of real and synthetic behaviours.

ACKNOWLEDGEMENTS

I would like to thank John Andreae and Brian Gaines for many stimulating discussions related to this work. The paper was written while I was an Erskine Fellow at the University of Canterbury, New Zealand, and the hospitality of the Department of Electrical Engineering there is gratefully acknowledged.

REFERENCES AND NOTES

1. G. J. Chaitin, "On the length of programs for computing finite binary sequences." *Journal of the Association for Computing Machinery*, **13,** No. 4, October 1966, pp. 547–569.
2. A. Nerode, "Linear automata transformations." *Proceedings of the American Mathematical Society*, **9,** 1958, pp. 541–544.
3. B. R. Gaines, "Behaviour/structure transformations under uncertainty." *International Journal of Man-Machine Studies*, **8,** No. 3, May 1976, pp. 337–365.
4. A. Gill, *Introduction to the Theory of Finite-state Machines.* McGraw Hill, New York, 1962.
5. P. H. Starke, *Abstrakte Automaten.* VEB Deutscher Verlag ver Wissenschaften, Berlin, 1969. Available in English translation from North-Holland, 1972.
6. A. Lempel and J. Ziv, "On the complexity of finite sequences." *IEEE Transactions on Information Theory*, *IT-22*, No. 1, January 1976, pp. 75–81.
7. R. McNaughton and S. Papert, *Counter-free Automat* MIT Press, 1971.
8. T. G. Evans, "Grammatical inference techniques in pa tern recognition." In: *Software Engineering, Volume* edited by J. T. Tou, Academic Press, New York, 197 pp. 183–202.
9. K. S. Fu and T. L. Booth, "Grammatical inferenc Introduction and survey—Part I." *IEEE Transactions Systems, Man, and Cybernetics*, **SMC 5,** No. 1, Janua 1975, pp. 95–111.
10. A. W. Biermann and J. A. Feldman, "On the synthesis finite-state machines from samples of their behaviour *IEEE Transactions on Computers*, **C-21,** No. 6, June 197 pp. 592–597.
11. I. H. Witten, "Learning to control sequential and no sequential environments." *Journal of Cybernetics a Information Science*, **1,** Spring-Summer-Fall 1977, pp. 9 111.
12. J. H. Andreae, *Thinking with the Teachable Machin* Academic Press, London, 1977.
13. B. R. Gaines, "On a danger in the assumption of ca sality." *IEEE Transactions on Systems, Man, a Cybernetics*, **SMC-6,** No. 1, January 1976, pp. 56–59.
14. Note that the null symbol ε is *not* the same as t empty string. Rather, it may be considered to be blank.
15. J. H. Andreae and J. G. Cleary, "A new mechanism f a brain." *International Journal of Man-Machine Studie* **8,** No. 1, January 1976, pp. 89–119.
16. F. R. A. Hopgood, *Compiling Techniques.* MacDona London, 1969.
17. $k-2$ unreachable leading-in states are characteristic canonical length-4 models. They can be avoided by p preceding the sequences with $k-2$ nulls before mode ing. The effect of this is to lose the initial part of behaviour from the model. However, if it will be r peated eventually this is only a temporary loss.

IAN H. WITTEN graduated with a B.A. in mathemati from the University of Cambridge, England; an M.Sc. in compu ing science from the University of Calgary, Canada, and a Ph.I in electrical engineering science from the University of Esse England. He was a Commonwealth Scholar at Calgary durin 1969–1970 and has been a lecturer in the Department Electrical Engineering Science at Essex University since then.

His research interests span the field of man-machine sy tems, and he has specialized in the fundamental problems machine learning; and in computational phonetics—the scienc of speech synthesis by computer. He is the author and co author of numerous publications, including several on th two-armed bandit problem.

During a sabbatical year in 1977, he worked at th University of Canterbury, New Zealand, on learning ma chines; at the University of Calgary on speech synthesis b computer, and at Bell-Northern Research, Montreal, o speech analysis.

t. J. General Systems
79, Vol. 5, pp. 13–16

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

HOMOMORPHISMS IN THE THEORY OF MODELLING

NORMAN Y. FOO
Department of Computer Science, University of Sydney, Sydney 2006, New South Wales, Australia

(*Received April 3, 1978; in final form September 11, 1978*)

Properties of dynamic systems preserved under homomorphisms are investigated. A result by Lyndon in algebraic logic is shown to apply to dynamic systems and characterises all first-order properties preserved. The justifying conditons of Zeigler are elucidated in this scheme and it is argued that there is no algorithm for discovering them. Higher order properties are briefly discussed.

INDEX TERMS Homomorphism, justifying conditions, decidability, Lyndon's theorem, simulation, model inference.

OMOMORPHISMS IN THE THEORY OF ODELLING

ne homomorphism approach to the theory of odelling has proved to be a very fruitful one. A cent book by Zeigler[10] has summarized the hievements of this approach in a language and a level aimed at a very wide and diverse adership. This paper is an attempt to draw the tention of systems theorists to a few problems hich are only partially resolved at the moment d should therefore deserve more investigation.

SYSTEMS AS ALGEBRAS

formal treatment of systems sufficient to en- mpass most of the requirements of modelling r digital computer simulation may be found in e reference cited above. We recall very briefly at part of the formalism necessary for this per. Basically a *system* is a quintuple (X, Y, Q, λ) where X is a set of admissible input gments, Y is a set of outputs, Q is a set of ates, $\delta: Q \times X \to Q$ is the state transition fun- on, and $\lambda: Q \to Y$ is the output function. Given o systems $S_1 = (X_1, Y_1, Q_1, \delta_1, \lambda_1)$ and S_2 $(X_2, Y_2, Q_2, \delta_2, \lambda_2)$ a *homomorphism* from S_1 S_2 is a triple (g, h, k) of maps $g: X_1 \to X_2$, $Q_1 \to Q_2$, $K: Y_1 \to Y_2$ such that $h\delta_1(q, x)$ $\delta_2(h(q), g(x))$ and $k\lambda_1(q) = \lambda_2(h(q))$ for each $\in Q_1$ and $x \in X_1$. This merely says that what ppens in system S_1 is mapped into system S_2 correctly. In fact the view is often taken that by asserting S_2 is a *valid* model of S_1, one is in fact asserting that S_2 is at least a homomorphic image of S_1.

Another way of regarding such a formalization of systems is to appeal to the notions of universal algebra. Such formalizations have been considered previously by Birkhoff and Lipson,[1] and Goguen, Thatcher, Wagner and Wright[6] among others, but we will specialize those developments here to a level sufficient for our needs. This has the advantage that there are extensively developed tools and results which could be immediately translated into the systems interpretation once an identification between the quintuple definition and algebras is made. The reader is referred to Cohn[2] for the notation and terminology we shall use. An *Ω-algebra* comprises a set Ω of *operators*, and a set A called the *carrier* on which these operators act. We claim that a *system* as defined above is an example of such an *algebra.* To see this the following identifications are made—the carrier of the (system) algebra is $Q \times Y$, and the operators are δ_x where

$$\delta_x : Q \times Y \to Q \times Y$$
$$(q, y) \to (\delta(q, x), \lambda(q))$$

i.e., δ_x actually comprises two functions δ and λ, and the second component in its domain is a dummy set. The identification may appear somewhat perverse but a justification really lies in the

usefulness of the observations in the sequel. A *system homomorphism* of system S_1 to system S_2 under this identification is in fact *algebra homomorphism* from algebra $Q_1 \times Y_1$ to algebra $Q_2 \times Y_2$. The carrier-to-carrier map required is simply the pair (h, k), and the operator association map takes $\delta_x = (\delta_1, \lambda_1)$ to $\delta_{g(x)} = (\delta_2, \lambda_2)$.

2. FIRST-ORDER SYSTEM PROPERTIES

Zeigler and his co-workers have raised interesting questions about various refinements of the preceding system homomorphism, the most basic of which appears to be this—if system S_2 is provably a homomorphic image of system S_1, under what conditions can we infer that a property of S_1 holds in S_2, or vice versa? A general framework for discussing this and related problems has been suggested by the author,[3] but it is only necessary here to retrieve a little of that suggested vocabulary. With regard to the system pair S_1 and S_2 above, the first posed question is called a *direct inference*, and the second a *converse inference* for the given property. In general, for an arbitrary property there is no reason to believe either inference is possible. However, by adding suitable conditions on S_1 or S_2, or both, these inferences may indeed become possible. The role of such conditions is made clear later but for the moment, following Zeigler, we call them *justifying conditions.*

Lyndon[7] in 1959 proved a theorem about algebras which partially answers the direct and converse inference problem for a large class of properties.

We recall informally some terminology necessary to understand this theorem. Exact definitions may be found in Cohn[2] and Yasuhara.[9] A *property* is formalised by writing down in some formal language (usually one which includes symbols for variables, functions, relations, quantifiers, and the logical connectives "AND", "OR" and "NOT" written respectively "$\wedge$", "$\vee$" and "$\neg$") a sentence which expresses that property. A property is *first-order* if the sentence which defines it is first-order, i.e., there is no quantification over functions or relations. Essentially, the first-order properties are those which are algebraic; thus, topological notions like convergence, continuity and compactness are not first-order. A sentence is *positive* if it does not contain the NOT connective. Lyndon's theorem asserts that if A is an algebra and P is a first-order property true of A and P has defining sentence α_p, then P is also true of *any* homomorphic image of A if and only if α_p is logically equivalent to a positive sentence. Without even formalizing a system theory to the extent of, say, Mesarovic and Takahara,[8] the reader can now verify by writing down the appropriate defining sentences for such system properties as *connectedness*, *linearity*, *controllability* and *time-invariance* that these are all positive in the above sense, and hence must be preserved in all homomorphic image systems if they hold in the original system. The contrapositive form of Lyndon's theorem is more interesting. It states that a first-order property is reflected back from a homomorphic image to pre-images if and only if the property is definable by a *negative* sentence, i.e., the negation of a positive one. Immediately, one obtains the *converse inference* result that the *negation* of the four aforementioned properties may be correctly inferred backwards from "models" to "real" systems if we are assured they lie in a homomorphic relation.

Asymptotic stability is definable as the existence of a fixed point in some suitable state space, and it may also be verified to be a positive property. It is common in simulation experiments to investigate through these experiments whether or not the "real" system is asymptotically stable. If the simulation model is believed to be a homomorphic image of the system, then the above discussion yields the following directive to the systems practitioner: *believe any instability results obtained but be skeptical if none can be found.* The more sophisticated systems practitioners, we feel, already understand directives like these intuitively.

3. JUSTIFYING CONDITIONS

We return to a discussion of Zeigler's *justifying conditions.* In the proof of Lyndon's theorem a rather stronger statement was shown to be true, viz, if Γ is the set of axioms of the algebra, then property P with defining sentence α_p is preserved in all homomorphic images of a system satisfying α_p is equivalent to some positive sentence *assuming the axioms.* The axioms delineate the system theory. For instance, in the theory of discrete-time discrete-state sequential systems Γ can be the collection of all the postulates appropriate to this theory suitably written in some

mal language. The essence of a paradigm vocated by Zeigler is to examine properties ich are not necessarily preserved under ho- morphisms, but which may be so preserved en sufficiently strong conditions are placed on systems concerned.[10] Such conditions justify transfer of assertions about systems to their dels and vice-versa, and hence are appropriately called justifying conditions. The role of ch conditions can be understood in our present ntext as follows. Let Γ be the axioms of a tem theory, and let α be a sentence which is e of one system S satisfying Γ, i.e., S is in the ss of systems being considered and α is true of Lyndon's theorem states that within Γ, α is e of all homomorphic images of S if and only here is some positive β equivalent under Γ to If this is not the case, but with the addition of urther axiom γ (not in Γ) α becomes equivalent some positive β, then γ is a *justifying condition* both the *system and its models*. For, using the tation in Cohn[2] or Yasuhara,[9] $\Gamma \cup \{\gamma\} \vdash \alpha \leftrightarrow \beta$ if d only if $\Gamma \vdash \gamma \supset (\alpha \leftrightarrow \beta)$ by the deduction theo- n and γ generally a sentence.

Our next observation provides evidence that search for justifying conditions is essentially a ative task. We do this by indicating that the of sentences that can function as justifying nditions is a recursively enumerable but not ursive. For a system theory with recursive set axioms Γ and some arbitrary sentence α noting a property of interest, let $J(\Gamma, \alpha) = \{\gamma | \Gamma \supset (\alpha \leftrightarrow \beta)$, γ a sentence and β a positive tence$\}$. We claim $J(\Gamma, \alpha)$ is recursively enu- rable. Since Γ is recursive, and presumably the es of deduction in the system theory are o recursive, the set of provable formulas from is recursively enumerable. Let these be listed $\mu_2, \ldots$ The set of all positive formulas is also ursively enumerable so let these be listed $\beta_2, \ldots$ Now all we have to do is to list cessively $\gamma \supset (\alpha \leftrightarrow \beta_1)$, $\gamma \supset (\alpha \leftrightarrow \beta_2), \ldots$ and use a -zag procedure to check this last list to see if formula in it is a theorem. This proves the im. However, $J(\Gamma, \alpha)$ cannot be recursive, for ndon showed that in general equivalence to a sitive sentence is undecidable, while it is clear t α is equivalent to some positive β under Γ if d only if any identically true sentence such as $\neg\mu$ is in $J(\Gamma, \alpha)$.

t is not unreasonable to attribute the usual ficulty encountered in looking for simple jus- ing conditions to the above fact.

Having found justifying conditions γ and γ' for a property defined by α within the system theory of Γ, γ' is a *weaker* condition than γ if $\Gamma \vdash \gamma \supset \gamma'$. Using the result above we can demonstrate the following assertions: (i) for γ and γ' in $J(\Gamma, \alpha)$ it is not decidable if $\Gamma \vdash \gamma \supset \gamma'$ (ii) for γ in $J(\Gamma, \alpha)$ and $\Gamma \vdash \gamma \supset \gamma'$ it is not decidable if γ' is in $J(\Gamma, \alpha)$. In either case the choice of an identically false sentence such as $\mu \wedge \neg\mu$ for γ reduces the recursiveness of $J(\Gamma, \alpha)$ to the questions for γ'. Therefore there does not exist any algorithm to tell in general if a given justifying condition is weaker than another.

Despite all these negative results it should be clear that an *efficient* semi-decision procedure for logical equivalence to positive sentences would be a great help to the modelling community. The author is not aware of the existence of any efficient procedure.

4. HIGHER ORDER SYSTEM PROPERTIES

Although the Lyndon's theorem provides a neat characterization of the direct and converse inference problem for systems in a homomorphic relation it does not apply to properties which cannot be expressed in a first-order language. The simplest instance of such a property is *finiteness* which is at least second-order (depending on the type of second-order language we admit). This is a well-known corollary of the compactness theorem for first order theories. A number of significant inference results invoke finiteness,[10] but Lyndon's theorem cannot be applied to obtain them. Other inference results of interest in modelling and simulation relate to *topological* properties like continuity which are essentially second or higher order. Examples of these are given in Foo.[4,5] There is, as yet, no simple characterization of higher order properties preserved under homomorphisms, not even when very strong topological conditions, e.g., the homomorphisms are both continuous and open, are invoked. The current strategy is not to proceed via logic or even model theory but to resort to techniques established in topology to obtain the higher order justifying conditions. Clearly, there is ample motivation for a sustained effort to resolve the questions in this area.

5. CONCLUSIONS

We have recalled how systems may be regarded as algebras, and hence justified the use of a

central result in algebraic logic by Lyndon to characterize all first-order system properties preserved under homorphisms and inverse-homomorphisms. The role of Zeigler's justifying conditions within this scheme was elucidated and we argued that the difficulty of obtaining such conditions could be attributed to their non-recursiveness. Finally, the limitations of the characterization were pointed out.

REFERENCES

1. G. Birkhoff and J. D. Lipson, "Heterogeneous Algebras." *Journal of Combinatorial Theory*, **8,** 1970, pp. 115–133.
2. P. Cohn, *Universal Algebra.* Harper and Row, New York, 1965.
3. N. Foo, *A Formal Theory of Modelling.* Unpublish manuscript, 1976.
4. N. Foo, "Stability Preservation under Homomorphism *IEEE Trans.*, **SMC–7,** No. 10, Oct. 1977, pp. 750–754.
5. N. Foo, "Closure Properties and Homomorphisms Time-Varying Systems" (to appear). *Mathemati Systems Theory*, **12,** 1978.
6. J. Goguen, J. Thatcher, E. Wagner and J. Wright, "Init Algebra Semantics and Continuous Algebras." *JACM*, **2** No. 1, Jan. 1977, pp. 68–95.
7. R. Lyndon, "Properties Preserved und Homomorphisms," *Pacific J. Math.*, **9,** 1959, pp. 143–15
8. M. Mesarovic and Y. Takahara, *General Syste Theory—Mathematical Foundations.* Academic Pre New York, 1974.
9. A. Yasuhara, Recursive Function Theory and Log *Academic Press*, New York, 1970.
10. B. Zeigler, *Theory of Modelling and Simulation.* Jo Wiley, New York, 1976.

NORMAN FOO is currently a lecturer in Computer Scien at the University of Sydney, Australia. He was Assista Professor at the School of Advanced Technology, SUN Binghamton from 1974 to 1975 and prior to that was research assistant with the Logic of Computers Grou University of Michigan. He obtained his Bachelors degree Electrical Engineering from the University of Canterbu New Zealand and his doctorate in Computer a Communication Sciences from the University of Michiga Ann Arbor. His interests are in systems theory, complexi and simulation.

. General Systems
Vol. 5, pp. 17-30

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

A GENERAL SYSTEMS THEORY OF PRODUCTIVITY†

RICHARD O. MASON

Associate Professor, Graduate School of Management, University of California, Los Angeles, California U.S.A. 90024

(*Received September 25, 1977; in final form February 6, 1978*)

Productivity is treated as a systems concept. It is argued that the causal texture of the system's environment dictates the relevant productivity construct and measurement. Based on Emery and Trist's taxonomy of environments, three concepts of productivity are defined: Process productivity, bounded productivity and systemic productivity. Each concept implies a different measure of productivity. Current economic and managerial literature is related to these concepts.

INDEX TERMS Productivity, environment, general systems theory.

TEMS CONCEPTS AND ODUCTIVITY

ductivity is a systems concept. It measures the of output a system has generated in relation ne input resources consumed during the pro-ion process. In the narrow sense, produc-y is a ration of outputs over inputs, out-input. In a broader sense, however, it con-s the overall efficiency of a system; and, it is broader aspect which we shall explore in this r. The rational system's manager generally eves efficiency by following three steps: (1) ulating the productivity of the existing pro-ive processes; (2) Comparing the result with native processes and; (3) Re-allocating re-ces to change those processes which are less ent or are wasteful. The route to efficiency, efore, starts with the calculation of produc-y. Query: On what basis are these calcu-ns to be made?

here are a staggering number of possible uctivity calculations. In order to establish a nale for selecting appropriate calculations, it seful to look at the properties of systems nselves and see what these properties imply in way of a meaningful basis for productivity ulations.

enerally speaking, a system is a set of co-nated components. It accepts inputs and engages in processes which yield outputs. According to Churchman[3,4] the study of systems involves five basic considerations.

1) The total system objectives, purposes and measures of performance.

2) The system's resources: the means used for performing the system's tasks and yielding its outputs.

3) The system's environment: the fixed constraints.

4) The components (elements) of the system including their activities, goals and measures of performance.

5) The system's management which acquires the resources and allocates them in order to achieve the system's objectives. Management encompasses the system's decision processes.

An understanding of the system's goals and objectives reveals which of the system's outputs are important for productivity analysis. Essentially the system's goals and objectives define the "functional class" of outputs which it is the purpose of the system to produce.[1,4]

This specifies the numerator of the productivity ratio. A thorough recording of the system's resources describes the means which may be employed in producing the output. The cataloging of resources influences the specification of the denominator in the productivity ratio; however, the basis upon which resources are evaluated depends on the nature of the system's environment. The more dynamic the environment the

ne author is indebted to the Editor, the Referees and to urton Swanson for valuable comments on an earlier n of this paper.

more relevant opportunity concepts are to the evaluation of resources.

In order to develop a systems concepts of productivity it is necessary to draw on a theory of systems' environments. For this purpose I have chosen the causal texture theory of Emery and Trist.[12] In the following section each of their four "ideal type" environments is summarized together with implications for productivity measurement.

ENVIRONMENTAL TYPE AND PRODUCTIVITY

Type 1: Placid, Randomized Environment

> In the simplest type, goals and noxiants are relatively unchanging in themselves and randomly distributed. A critical property from the organization's viewpoint is that there is no difference between tactics and strategy, and organizations can exist adaptively as single, and indeed quite small, units.

The implication of this environment for a system is that it makes it stable and relatively unchanging. The productive function may be characterized as a stable channel through which inputs and outputs flow. An appropriate measure of productivity is a measure of the efficiency of the system for converting inputs into outputs. Thus, the traditional measure, output/input, is applicable I call this a measure of *process* or *flow productivity.*

Type 2: Placid, Clustered Environments

> The next type is also static, but goals and noxiants are not randomly distributed; they hang together in certain ways. Now the need arises for strategy as distinct from tactics. Under these conditions organizations grow in size, becoming hierarchical and tending towards centralized control and coordination.

The clustering aspects of this type of environment give rise to a concentration of resources and the development of a "distinctive competence" which limits the system's output productive capacity. Yet, the placid nature of the environment means that the constraints on the system are not changing rapidly. What a system *could* do with its resources rather than what it *is* doing is the relevant concern. Failure to utilize the system's full or most efficient capacity results in waste and inefficiency. It represents a loss of the system's opportunities and is indicative of an improper concentration of resources. The relevant measure becomes a comparison of actual with potential output. Because it focuses on t constraints imposed by the system's particul concentration of resources, I call this measu *bounded productivity.* Bounded productivity similar to Stafford Beer's[2] concept of "produ tivity" which he defines as the ratio of a syste actual output ("actuality") to its achievable ou put ("capability").

Type 3: Disturbed, Reactive Environments

> The third type is dynamic rather than static. It consists c clustered environment in which there is more than system of the same kind, i.e., the objects of one orga zation are the same as, or relevant to, others like it. S competitors seek to improve their own chances by hind ing each other, each knowing the others are playing same game. Between strategy and tactics there emerges intermediate type of organizational response—what milit theorists refer to as operations. Control becomes m decentralized to allow these to be conducted. On the ot hand, stability may require a certain coming-to-terms tween competitors.

The reactive nature of this third type of e vironment means that there are competitors a that the system must consider that its knowled and production capacity can be acquired others. The fact that the environment is disturb and changing means that there always are n opportunities being created in the system's vironment which permit it to restructure its and to respond to the actions of competitors a others. The critical facility for long-run survival the system's manager's ability to reach out in the environment, acquire new resources, alter t productive function, revise the flow of inputs a outputs and to re-configure a new system wi new capacity. Failure to configure the best p ssible system results in lost opportunities and waste and inefficiency. The relevant measure comes a comparison between the existing syste and the best possible system. Since it involves concepts of investment and the re-configuring the system, I call this a measure of *syste productivity.* Systemic productivity is related Beer's[2] concept of "performance" which he d fines as the ratio of actual output ("actuality") what the system *ought* to be doing by developi its resources and removing constrai ("potentiality").

Type 4: Turbulent Fields

> The fourth type is dynamic in a second respect, dynamic properties arising not simply from the interact of identifiable component systems but from the field it

he 'ground'). The turbulence results from the complexity nd multiple character of the causal interconnections. ndividual organizations, however large, cannot adapt suc-ssfully simply through their direct interactions. An exam-ation is made of the enhanced importance of values, garded as a basic response to persisting areas of relevant ncertainty, as providing a control mechanism, when com-only held by all members in a field.

At this fourth level of environment the concept efficiency is superceded by the concepts of vival and effectiveness. The functional class of puts is no longer well defined. Thus, the vant managerial issue becomes the choosing the "domains"[25] in which the system will rate. In this context resources may be radi-y re-oriented to produce new outputs; and, eed, many resources may be kept idle, in rve for some yet to be identified new domain. *value* of efficiency is important; but, a stable asure of productivity cannot be defined.

Thus, there are three idealized environments in ch productivity measures are relevant. nerally as the system becomes more open and amic with respect to its environment, the ropriate productivity measure changes from a ely internal measure to a comparatively exter- measure. In highly complex environments the asures appropriate for a more stable environ-nt may be usefully calculated. However, they st be interpreted in the context of the pro-tivity measure defined for the environment in ch the system finds itself.

With this background on the role of systems ory for defining productivity measures, we n next to an explication of process, bounded systemic productivity.

OCESS PRODUCTIVITY

cess productivity is the concept of produc-y most generally used in economic theory. It neasured by calculating the ratio of what is duced by the system (output) to what the em consumes in producing it (resource input). s notion of productivity leads to a deceptively ple formula for process productivity, O/I. iculties in applying this formula result when s of output and input must be defined and s of measure specified. This has led to much fusion in the economic literature. To sort e problems out it is useful to view process ductivity by means of a systems model such hown in Figure 1.

The system's manager seeks to maintain both the technical efficiency of the system and the market and environmental viability of the system. For the former the manager is concerned with the physical relationships between input and output. The latter is concerned with the monetary relationship between expenses (outgo) and sales (income or revenue).

a. *Physical Process Productivity*

To measure physical process productivity outputs are expressed in physical, tangible units of goods or services. Number of items, tons, kilowatt hours, gallons, square feet, etc. are typical units of measure used. Inputs are expressed in terms of actual amount of resources consumed in producing the output and are usually classified into labor, materials, energy, capital and other factors of production.

From a system's point of view the relevant measure of productivity is a ratio of output to *all* associated inputs. This is generally referred to as a "total factor productivity" measure and it reflects the overall efficiency of the systems processes for converting inputs into outputs. The association of inputs to outputs is accomplished by means of an activity analysis. A full scale activity analysis of a multiple technology would yield an activity matrix A. The A matrix takes the form:

$$\begin{array}{llll} I_{1,1}; & I_{2,1} & \cdots & I_{n,1}; \quad O_1 \\ I_{1,2}; & I_{2,2} & \cdots & I_{n,2}; \quad O_2 \\ I_{1,m}; & I_{2,m} & \cdots & I_{n,m}; \quad O_m \end{array}$$

A total factor productivity measure reflects input as a function of the amounts of the nth input resources consumed in producing a particular output. Consequently, process productivity measures deal with only the rows of this A matrix. As will be discussed in a subsequent section, bounded productivity measures permit comparisons to be made among the rows.

It is important to point out that many changes in the system's conversion processes may explain the changes observed in a total factor productivity measure. New technology, re-organization, new management, new materials, changes in labor skills and many other factors may account for increases or decreases in the ratio.

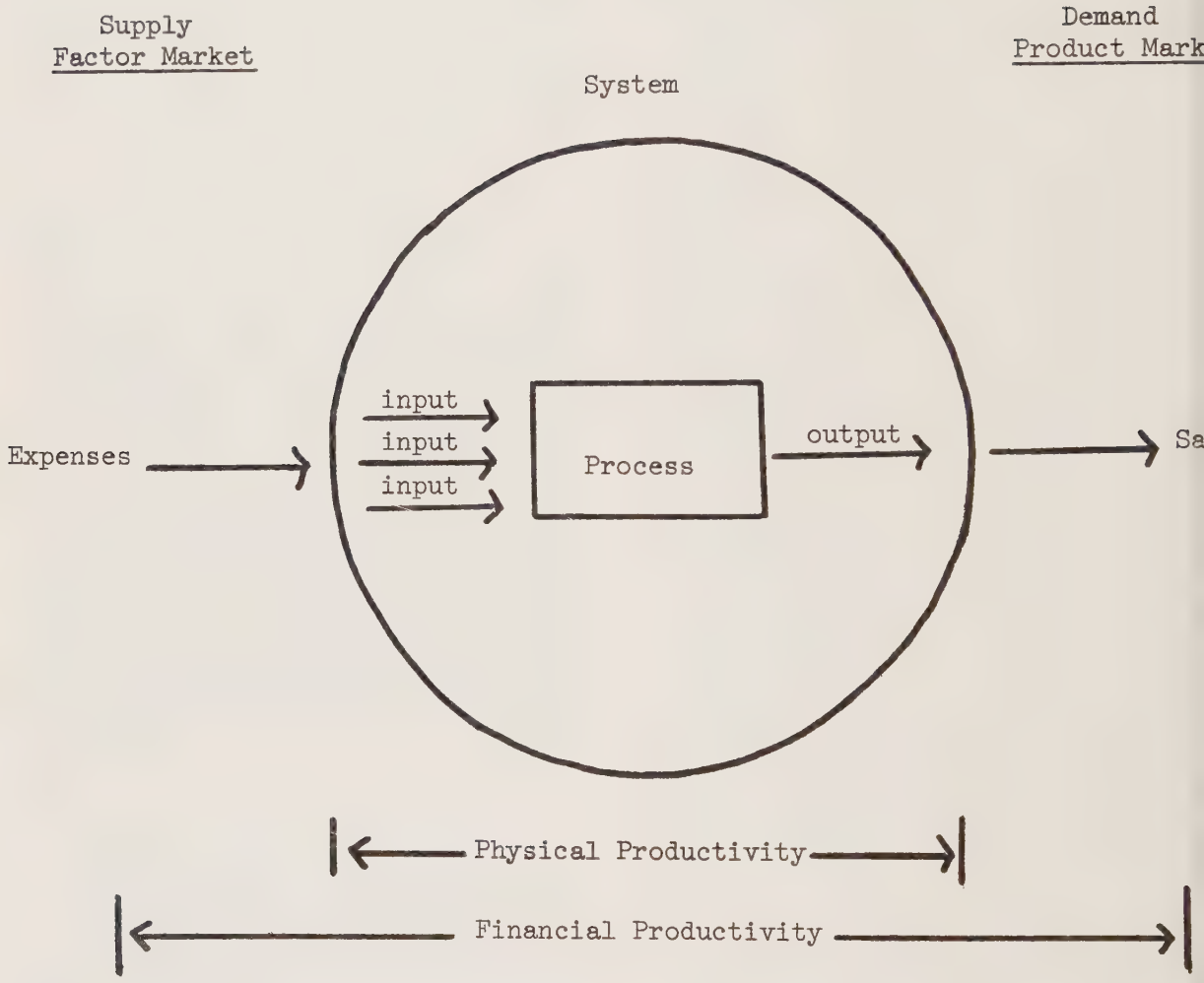

Supply
Factor Market
Demand
Product Mark
System
input
input
input
Expenses
Process
output
Sa
Physical Productivity
Financial Productivity

FIGURE 1

Partial productivity measures help identify me of the sources of productivity change and e easier than total measures to operationalize. partial productivity measure is the ratio of tput to a single or limited set of elements of put. It takes the form,

$$\frac{O^b}{I_i^b}$$

r a base year, where I_i^b represents the ith ement(s) of input. An index of the form

$$\frac{O^t/I_i^t}{O^b/I_i^b}$$

n also be calculated for any comparative per- d t.

Output per man hour is the most frequently ed partial measure. In fact it is so common ace that many people define "productivity" clusively as output per man hour. For many stems this is a reasonable measure to use since bor is a major input factor and because labor iciency has been an historical focal point for anagerial concern. A partial measure such as tput per man hour also reflects changes in the rformance of the system as a whole including e overall efficiency with which resources are ed, the average quality of labor, the quality and ailability of materials and the methods of orga- zation and work flow.

There is a deep social value underlying the use a physical process productivity. In essence it is measure of society's ability to overcome the rces of nature. More output or use-value per it of input is a measure of a system's (and, nce, man's) increasing ability to create "plenty" d abundance. This is the concern of the pro- cer, the engineer, the technologist, and the anager. There is a long history in economic erature for this concern. The primary focus of s concern has been on labor input.

Adam Smith[31] argued that "every man is rich poor according to the degree in which he can ord to enjoy the necessaries, conveniences, and ausements of human life." This led him to place e value of a commodity as "equal to the antity of labour which it enables him to pur- ase or command." Outputs and inputs in this w should be measured in terms of their exchange or market value not in physical terms. In this view a system's outputs and inputs are valued in terms of their "scarcity" value.

David Ricardo, however, disagreed with Adam Smith. The fundamental system's problem was to secure abundance from nature. The proper value is "use" value. The value of an output is not inherent in the quantity of labor purchased or commanded from others in terms of its scarcity; but, rather in the physical quantity of labor which was consumed in the process of producing the output. Ricardo thereby gave rise to the labor theory of value, a theory in which output per man-hour is a central notion. Karl Marx was to extend this notion.

The labor and use theories of value are also the basis of much of contemporary industrial engineering. Time and motion study and scientific management, as developed by Frederick W. Taylor, are notable examples. Taylor sought to reduce the gap between what workers actually turned out and what they might comfortably turn out. His approach was to study the processes of work in detail. The amount of time required to accomplish a task using one method was compared to the amount required using other methods until the most efficient method was found. In his studies Taylor identified three forms of measurement which were essential to scientific management: a measure of the quantity of physical output (e.g., pounds, tons, bushels), a measure of the quality of physical output (e.g., grades, errors), and the productivity measure of man-hours per unit of output required to produce the output. Taylor's objective was to solve the engineering problem of improving a worker's capacity by discovering better methods. He also sought to solve the psychological problem of improving his willingness to work by developing incentive systems. The goal of these solutions was to increase output per man hour. This tradition, frequently called Taylorism, has continued through much of industrial engineering and management science.

Physical process productivity measures do ignore scarcity, however. A system's manager is also concerned with the financial viability of the system in a market and monetary economy. In a monetarized economy, prices are a measure of the relative scarcity of goods and factors. Indeed, since John Stuart Mill first formulated the monetary measure of value, most of the focus of economics has been on the monetary or price

side of systems. Most systems are compared in terms of their relative financial performance.

Productivity analysts have used monetary data extensively for their studies. There are several reasons for this. One is that some analysts firmly believe that a measure which relates sales or income to expenses is the relevant measure of efficiency for a system. Another prevalent reason is the availability of dollar data for most elements of production. Since the dollar is a common unit of measure among goods and factors, prices are often used as a surrogate for physical relationships. A third explanation, of course, is that they have confused the use value premises, which underly physical relationships, with the scarcity and exchange value premises, which underly price-market relationships.

b. *Financial Process Productivity*

For a business enterprise, profit is the purest measure of management's ability to operate efficiently in the market place. It is generally calculated as Sales (Income or Revenue) minus Expenses, that is, Income − Outgo. Sales or Income are normally used as the output measure in the private sector. Revenue is normally used in the public and not-for-profit sectors. This relation may be expressed as a ratio S^b/E^b. The ratio is greater than one when the firm is profitable or the non-profit organization has an excess of revenue over expenses. The ratio is less than one when outgo exceeds income. Trends can then be established by comparing a base period ratio with subsequent periods by means of an index.

While a productivity type measure of the form S/E can be calculated, the more typical measure is the profitability rate, $S-E/S$. This is just one of several financial ratios which are used to measure the financial productivity of an organization. Return on investment or return on assets,

$$\frac{\text{Profit}}{\text{Investment}} = \frac{S-E}{A},$$

and the capital turnover ratio,

$$\frac{\text{Revenue}}{\text{Investment}} = \frac{S}{A},$$

and other variations of investment based ratios are also frequently employed for financial analysis.

These financial measures—profit rate, return o investment and capital turnover—are the pure reflection of the organization's ability to co with problems of scarcity in a monetarized ar market economy in which a monetary unit (suc as dollars) is employed as a medium of exchang These financial measures are not normally co sidered to be productivity measures, but rath measures of financial performance. However, th form one end of a continuum of measures ru ning from pure financial measures on one end pure physical measures of productivity on t other end. The middle of this continuum contai many "mixed" measures in which prices ar financial inputs and outputs are co-mingled wi measures of physical inputs and outputs.

Two frequently used measures which mix fi ancial and physical dimensions are cost per un of output and revenue per man-hour. Cost p unit of output is the fundamental notion of co accounting. It is the result of dividing all releva financial input costs by the number of units physical output,

$$\frac{P_1 I_1 + P_2 I_2 \ldots P_n I_n}{O}.$$

P_i = the factor price for the ith input, I_i = t quantity of the ith input, and O = the physic quantity of output. Cost per unit is influenc not only by the technical efficiency of t organization's conversion processes but also the supply forces affecting its input factors a the organization's skills and relative power in t factor input market. Consequently, cost per un is a measure of how well the organization h been able to acquire scarce resources from environment and convert them to physic output.

Sales or Revenue per man-hour is the conver of cost per unit. It is calculated by dividing sal by man hours $S/I_e = O \cdot M/I_e$ (where M = the pr duct market price of output O and I_e is t physical amount of labor input). (Actually, measure of sales per total factor input can calculated, but its use is rare. Some firms use related measure, however, in which the total sal or revenue for a product are divided by t direct, traceable costs of producing the produc Many industrial firms use the measure sales (revenue) per man-hour as indication of the overall productivity. One reason given for its u is that most employees and managers understa

t and because its behavior is believed to be orrelated with purer financial and physical meaures in the range in which the firm is operating. The partial financial measure Sales (or Revenue) er dollar of labor expense (or, sometimes, just irect labor expense) is used for the same purose. Sales per man-hour is a measure of the rganization's ability to employ its physical inuts to satisfy demand and scarcity requirements n its product market. It is influenced not only by he technical efficiency of the organization, but lso by the social demand for its output, the elative scarcity of the output, and the rganization's skill and relative power in the utput product market.

Some productivity analysts have attempted to educe some of the effects of scarcity factors on nancial measures by using "value added" rather han sales as a financial output measure. Value dded is the market price of the output minus he cost of materials purchased from others. Gross value added reflects the cost of labor, axes, profit, interest, rent and capital depreiation. Net value added excludes capital depreiation. Value added per man-hour has the adantage of not being distorted substantially by he extent to which a system processes its output hrough all stages of production from basic raw naterials to finished product. Hence, it is useful or comparing productivity between organiations with different degrees of plant integration. Value added per man-hour is a measure of the rganization's ability to apply its physical labor o the raw and semi-processed materials it acuires and to add scarcity or market value to nem. It is frequently used in industry level roductivity studies to compare one firm with nother.

One method for moving financial measures loser to physical measures is to employ price eflators. Constant dollar sales and expense fiures may be obtained by applying a price index vhich reflects the prices of the particular outputs nd inputs the system handles. Value added gures require a "double deflation" for both the alue of output and also for the cost of materials. Constant-dollar values may also be obtained diectly by multiplying the number of units of utput or of input by a unit price calculated for a ase period.

The use of price deflators reduces the effect of rice changes on productivity measures. owever, it is not without problems. There is a technical difficulty in estimating constant dollars because the appropriate or ideal price index for a given system usually is not available. There is also a conceptual problem in their use as physical measures. Under ideal circumstances they still reflect the relative financial scarcity in the system's market environment during the base period. Despite these problems, price adjusted data is commonly used. For many systems the requisite financial data is more readily available than the physical data and, especially for large and highly complex systems, constant-dollar productivity is most likely the closest approximation to physical productivity that can be obtained.

A recent paper by Craig and Harris[6] develops one of the most comprehensive approaches to constant-dollar financial productivity in the literature. The unique feature of the Craig and Harris model is the treatment of capital inputs. The authors reject measures of the physical input of capital (such as depreciation which is, admittedly, a poor measure of physical contribution) in favor of a service value of capital concept. They then turn to the financial market to estimate the service value input. The model assumes that the firm has a leasing subsidiary which purchases its land, buildings, equipment, cash and other assets. The capital input defined to be the payment made by the firm to the leasing subsidiary. The payment is calculated as an annuity on the basis of the base year cost of the asset, its assumed productive life and on the desired rate of return to the lessor. The desired rate of return is considered to be the relevant cost of capital to the firm.[16] A weighted average of the percentage elements of the firm's capital structure multiplied by each source's cost factor is suggested for estimating the cost of capital.

The Craig and Harris model has much to commend it; but, it is nevertheless a mixture of physical and financial measures. In addition to the problems of applying price level deflators its treatment of capital is specifically financial in nature. Cost of capital is a monetary measure of the relative scarcity of financial resources. Consequently, it reflects the firm's skills and power in the capital markets and investor expectations about the firm's earning power.

It is not surprising that financial and physical measures become intermixed in measuring productivity performance. They are inherently dialectical in nature. Scarcity considerations of supply and demand occur as man exchanges

goods and services with his fellow men. Efficiency considerations occur as man works against nature to increase the abundance of goods and services. Thus, the concepts are at once distinctively different and deeply interwoven. Scarcity is reduced by efficiency. Efficiency is directed and motivated by scarcity. And, hence, our financial measures of scarcity and our physical measures of efficiency in the end are bound to be interrelated.

In constructing a measurement system for the process productivity of a producing system we need not resolve this dialectic. We need only to understand it. In practice both financial and physical measures and their composites should be used. They each tell us something different about the productivity of the system and its ability to survive in its environment. Also, for reasons discussed above, both partial and total measures are appropriate. Consequently, process productivity is probably best measured by a "family" of productivity measures. Such a family of measures, of course, can be quite large. In the case of a simple production system with one output and n inputs there are 2^n-1 total and partial measures of physical productivity. Since there is an equal number of measures of financial productivity, of physical output to financial input and of financial output to physical input, a total of $4(2^n-1)$ measures are possible for each output.

As we have seen, productivity analysis often requires that we distinguish outputs by the different technological processes which produce them and that inputs be categorized into classes according to their contribution to the productive process. Consequently, for any but the most macro analysis a substantial number of theoretical measures is possible. It behoves the systems manager, therefore, to develop a strategy for choosing the subset of measures which will be most useful in dealing with scarcity and efficiency issues.

There is substantial literature on the subject of what is here called process productivity. An introduction to the topic is found in Fabricant's *A Primer on Productivity*.[13] Application to industrial firms is found in Kendrick and Creamer,[23] Greenberg,[19] and Eilon, Gold and Soesan.[11] Melman[26,27] has done related studies. Process type productivity measures for industry and the national economy are given in Kendrick[22] and in Denison.[8,9,10] Bela Gold[17,18] has dealt explicitly with the difference between financial and physical measures of productivity.

With a few exceptions (to be discussed in the next section), the measures in this literature are all process productivity measures as defined at the beginning of this section. The output to input ratios they contain reflect the system's ability to cope with scarcity and efficiency. As the ratio of output per unit of input increases, the system's power over nature and over other systems is improving. But, process productivity only tells the system manager how well the system *is* doing not how well it *could* do. A system which has accumulated resources in a placid, clustered environment must be concerned with how well it could do. Bounded productivity addresses this question of potential and deals explicitly with the system's performance in relation to its capacity and capability.

BOUNDED PRODUCTIVITY

Bounded productivity describes how well the system is doing compared with its potential given its *existing* resources and technological configurations. It is measured both in relation to output and to input.

Bounded output productivity is the ratio: actual output *divided by* the maximal output that a system with its particular concentration and configuration of resources (capital, labor, supplies, knowledge, etc.) could produce.

Bounded input productivity is the ratio: minimal input resources that should be required by a system of this configuration to produce a given level of output *divided by* the actual input resources consumed to produce that output.

In traditional economic analysis bounded productivity relates productive performance to that which would have occurred with the "frontier production function." A system's frontier production function expresses the maximal output obtainable from the system's particular combination of factors of production (resource inputs) at the existing state of technical knowledge.

There are several approaches to estimating a system's frontier production function output so that it may be used as the denominator in the boundary productivity ratio. These include: (a) determining the "engineered" production function; (b) estimating relationships on the basis of historical input and output flows; and (c) es-

nating relationships on the basis of external luencing or modifying factors.

The most natural method is to use the oretical function specified by the system's de- ners and engineers. It is possible to define such "engineered" production function by consult- current catalogue information and technical alysis of the flow of work to determine the tem's achievable output rate.

As an illustration of this approach, consider a tem consisting of a typist, a manual type- iter, paper, etc., the purpose of which is to oduce correctly typed pages as output. Suppose system is currently producing at a rate of 00 words per man-hour. Three thousand wo- per man-hour is its "partial" rate of process oductivity. Suppose that, according to technical ecifications of the equipment and analysis of ndard labor times, the system is designed to oduce at a rate of 7,200 words per man-hour. en its bounded productivity would be $\frac{3000}{7200}$ or 4167. Alternatively, since the world record (per- med on an Underwood Standard manual ma- ne) is about 8,820 words per man-hour, one ght calculate the lower level bounded pro- ctivity as $\frac{3000}{8820}$ or 0.3401. Increases in either of se ratios indicate that the system's resources being more fully utilized and, hence, the tem is more productive.

For many systems, especially in complex orga- ations, it is difficult to arrive at an agreed on designed, standard or maximal output. This particularly true in systems which involve bstantial administrative or indirect inputs, rning behavior, or which require co-ordinated er-personal relations. When this is the case a ximum can be estimated empirically. Such an pirical production function has the advantage t at least some of its points have been obser- d as outputs of real systems.

One approach for deriving an empirical pro- ction function for a system is to collect statis- s on inputs and outputs for a group of similar tems (say, all of the firms in an industry). This ta is used to infer an average production action. Normally, this will be an appropriately ighted mean of the production function as- ciated with each of the various technologies ed by the group of comparable systems. If the tem's analyst is willing to assume a linear, mogenous production function, then the input d output data can be processed using ordinary multiple regression methods to yield an average production function.

An alternative, but related, approach for deriving an empirical production function is that used by M. J. Farrell.[14] Farrell's method estimates the lower bound of the system's production function. The method essentially involves estimating a unit isoquant (points representing the minimum quantities of the resource factors of production required to produce one unit of output with varying factor proportions). An envelope is fit to the data from below where each data element consists of resource input data divided by the total output. The production function is then found by solving a series of simultaneous equations. A similar result can be achieved, however, with regression techniques by forcing the residuals to be positive.

Farrell's approach and also the "best practice" methods suggested by Anne Grosse[20] are based on relations between internal factors and output. Sometimes these relationships are difficult to establish. This is especially true in information producing and administrative areas. This difficulty has led some organizations to focus on estimating resource input standards as a function of some environmental parameters of the system. IBM, for example, has found that standard input resource levels for secretaries are more reliably estimated as a function of the number of personnel in the plant, than as a function of the number of reports, amount of correspondence, etc. produced. (IBM currently estimates some 110 administrative input standards on the basis of some 28 external environmental variables they call "modifiers" in their system.)

Standards, whether derived empirically or through engineering analyses, form the basis of a popular form of bounded productivity measure. This is the relation between standard hours or "earned" hours and the actual hours of labor input. Many organizations have calculated a standard for hours (and also for materials) per unit of output. They subsequently calculate the ratio:

$$\frac{\text{Standard input per unit of output } x \text{ units of output}}{\text{input actually consumed}}$$

This ratio is a bounded input productivity measure.

Systems theory also suggests an approach for

improving productivity as measured by the bounded concept. At any point in time the system's output is limited by just those constraints which are "binding." In the event that a system is achieving less than its maximum, that is, whenever a system's bounded output productivity,

$$\frac{\text{actual output}}{\text{maximal output}},$$

or its bounded input productivity,

$$\frac{\text{minimum input}}{\text{actual input}},$$

is less than one, then some unspecified constraints must be operating. Discovering these phantom constraints and releasing them will aid the system's management in increasing the system's productivity.

The institutional economist, John R. Commons, refers to these binding constraints as "limiting factors" and relates them to "complementary factors." Complementary factors are similar to resources in systems terms. In Commons' words:

> The limiting factor is the one whose control, in the right form, at the right place and time, will set the complementary factors at work to bring about the results intended. A very little potash, if that is the limiting factor, will multiply the grain yield from perhaps five bushels to twenty bushels per acre. The sagacious mechanic is the one who busies himself with control of the limiting factor, knowing that the complementary factors will work out the results intended. The flutterbudget wastes his time on the complementary factors." (pp. 628–629)[5]

In the typing example discussed earlier, it is possible that the factor limiting a typing system to 3000 words per hour when its achievable rate is 7200 words per hour is the skill of the typist. Hence, additional training may improve this limiting factor. Of course, there are many other possibilities. The output could be limited by the motivation of the typist or the implied or negotiated work rules which control the pace of work. The repair of the machine or the kind of paper, or the organization or the flow of work may constitute or contribute to the limiting factor. Or, it is also possible that the typist is not as productive as possible because some effort is devoted to complying with government regulations. Whatever their source, identifying the limiting factors is important in productivity analysis.

John Commons correctly points out that in a dynamic or on-going system the limiting and complementary factors are continually changing place. Thus, an effective management is one that pays attention, sequentially, to the active limiting factors. In the language of management planning releasing (i.e., converting to complementary) a limiting factor becomes a management *objective* Consequently, the sequential attention to objectives which Cyert and March[7] claim describe managerial behavior is also rational as a means of securing improved productivity in a placid clustered environment. The vast literature on management by objectives is also relevant in this regard.

Some recent work by Charles Kriebel and colleagues[24] contains an analytical method for searching for limiting factors, especially those of a technical nature. Using a form of activity analysis, Kriebel relates a vector of output rates, Y_j's for services j to a constraint vector X of x_i's Each x_i's represents the capacity service rate of input resource category i measured in resource processing units per unit time. For a given system technology, A, a standard product output mix $y^0 = (y_1, y_2, \ldots, y_n)$ can be specified as:

$$Ay^0 = x^0 \leqq X$$

The utilization of resources when the input employed are x^0 can be evaluated using this model. λ^* is defined as the maximal factor by which the amount employed of all resources for the product mix y^0 can be increased without violating the capacity constraints. The scalar λ^* is the solution to:

$$\lambda^* = \text{Max}\,\lambda$$
$$\text{subject to } \lambda^* x^0 \leqq X$$

If at least one of the resources (the limiting factor) is fully utilized $\lambda^* = 1$ otherwise $\lambda^* \geqq 1$ λ^{*-1} then becomes a bounded productivity measure indicating the level of resource utilization in the system. The method appears to be sound whenever the A's and X's refer to technical production systems. The method could be extended however to deal with less tangible resources and constraints such as motivational and regulatory factors.

The search for the maximal output or minimal input for a system may also be facilitated by

ıear programming. There is a close connection tween the activity analysis developed by oopmans and the simplex method for linear ogramming developed by George B. Dantzig. s described briefly in the preceding section, the chnical knowledge about a system may be presented in an activity analysis model and mmarized by a matrix A. When this model is nfronted with a given stock of factors of pro- ıction and a specific objective to be maximized g., output) or minimized (e.g., input), a linear ogram may be used to determine the system's ɔtimal output or input levels.

Bounded productivity as described here is im- ɔrtant because it serves as a guide to manage- ent. It tells the manager how far actual pro- ıction is from what is achievable with the stem's given resources. It spurs the search for niting factors. Hence, it provides valuable infor- ation for the practical and rational decision- aker. Measures similar to the bounded pro- ıctivity measures developed here are found in rious parts of the literature. Beer[2] has de- loped a series of measures of achievement in nich his measure of "productivity" is similar to e bounded measure developed here. In ad- tion, the rate of return on investment analysis DuPont and General Motors[21] relates output capacity as one of the explanatory variables. ela Gold[18] has linked ROI to more traditional oductivity analysis through the equation:

$$\frac{\text{Output}}{\text{Total Investment}} = \frac{\text{Output}}{\text{Capacity}} \times \frac{\text{Capacity}}{\text{Fixed Investment}} \times \frac{\text{Fixed Investment}}{\text{Total Investment}}$$

veral studies of productive systems have used a unded measure as their effectiveness criterion. or example in the Ahmedabad Experiments, A. Rice[28] employed engineering techniques to timate productivity as actual output divided by gineered potential output. M. J. Farrell's orig- al work has been extended by Farrell and eldhouse[15] and Wesley Seitz[29,30] Anne rosse's "best practices" method[20] is an alter- tive to that of Farrell.

Both process productivity and bounded pro- ctivity take the existing system as given. They e measures of how well the system is doing th the resources available to it. But has the system acquired all of the resources it might? Systemic productivity deals with this query.

SYSTEMIC PRODUCTIVITY

Systemic productivity questions the design of the system itself: "Is it living up to its maximum feasible potential?" "Is it doing what it *ought* to do to develop its resources and to pursue opportunities?" In a disturbed, reactive environment these are the salient questions. Strategic planning is the information base for answering these questions. Systemic productivity is the measure. Investment is the action orientation.

Systemic productivity describes how well the system is doing compared with its potential as measured by the performance of any feasible, designable system of the same functional class. A system of the same functional class is a system whose actual or potential output will achieve the same end or purpose in the system's environment. Presumably, the system's manager could have re-configured the system into any of those of the same functional class that were reasonably available in the system's environment. Consequently, systemic productivity is a measure of opportunities lost because of a failure to invest and to engage in systems evolution and redesign. As with bounded productivity it is measured both in relation to output and to input.

Systemic output productivity is the ratio: Actual output *divided by* the maximal output that could be produced by the best alternative, feasible, designable and importable system of the same functional class.

Systemic input productivity is the ratio: Miminal input resources that are required to produce a given level of output by the best alternative, feasible, designable and importable system of the same functional class *divided by* the actual input resources consumed to produce that level of output.

To return to the illustration of the typing system, the actual rate of 3000 words per man-hour using a manual typewriter was related to a design standard (7200) and a world record (8820) for manual machines in order to obtain the bounded productivity measure. The system, however, could be redesigned using an electric typewriter. Electric typewritten pages are of the same functional class (and, in this case, also of the same morphological class) as manually typewrit-

ten pages. However, the official world record on an electric machine is 9316 words per man-hour, recorded using an IBM electric. Actually, rates approaching 12,960 have been observed. This would suggest that one basis for calculating systemic productivity would be to relate the actual rate to this potential rate. The systemic productivity then would be

$$\frac{3000}{9316} = 0.3220 \quad \text{or} \quad \frac{3000}{12,960} = 0.2315 .$$

The investment decision to exchange a manual typewriter for an electric typewriter is a relatively minor system's redesign. Computerized word processing represents a more dramatic reconfiguration within the same functional class. While it is still limited to some form of manual key entry for original data, a word processing system can store, format and organize data so that the net effective output rates can be much higher. For example, it may be possible on current IBM equipment, to achieve as high as 45,000 words per man-hour. The systemic productivity of the manual typing system then would be

$$\frac{3000}{45,000} = 0.0667 .$$

Systemic productivity is an indicator of the long-run survival and viability of the system. It may be applied to work stations, organizations, industries and entire economies, alike. It's application raises questions the other measures miss. For example, it is possible that an industry, such as the U.S. steel industry, has process and bounded productivity measures which are within historically acceptable levels. That is, the industry may be relatively efficient given the turn-of-the-century technology many firms employ. However, in light of the development of new technology and methods in Europe and Japan, the industry's systemic productivity may be quite low. This low ratio is a signal to the industry's management that its long-range plans and its investment policy ought to be re-evaluated. The firm is in a disturbed environment and eventually its competitors may out-perform it.

Low rates (i.e., below one) of bounded or systemic productivity do not necessarily imply that immediate action should be taken to close the gap. The measure is a signal. It alerts management to conditions which should be understood and explained. For example, a low *bounded* productivity measure calculated on the basis of machine specifications may also be the result of a previously unconsidered constraint such as work rules negotiated in collective bargaining or work conditions imposed by compliance with an OSHA safety rule. Several judgments are possible in this case: (a) accept the imposed rates as binding and scale down the rest of the system to match; (b) attempt to negotiate a new standard; (c) continue to operate the system as is; or (d) seek ways to reconfigure the system within the confines of these constraints. The appropriate action would be determined by the result of a thorough productivity review.

Low *systemic* productivity measures suggest other options. Options for improvement include total redesign of the system, acquisitions of new equipment, and substantial retraining or hiring programs. In short, low systemic productivity is an indication that new investment is required. In terms of managerial action, there is a special relationship between bounded and systemic productivity. Bounded productivity focuses attention within the system to look for solutions; systemic productivity focuses attention without. However, if bounded productivity is low, then the system's management must be very careful in adopting systemic changes such as new equipment or new designs as a solution, unless these new solutions deal specifically with the limiting factor.

This is not unlike the story of the farmer who told the agriculture extension agent, "Sucks, I'm not farming as well as I know how now!" He was essentially saying, "Don't bother me with systemic changes. I'm not working up to my bounded productivity level yet." Consequently the first order of business for the system's manager is to see that the system's available physical, physiological and psychic energy is freed to do work and that this energy is appropriately directed. This requires the manager to employ a wide range of skills, including organizational development skills.

The difference between bounded and systemic productivity is still relevant however. It is a guide for new investment and the importing of new energy for the system. If a manual typing system has a bounded output rate of 7200 words per man-hour and a systemic output rate of 9316, or of 45,000 words per man-hour, then a potential of between 2116 or 37,800 additional words per

man-hour could be realized by an investment in an electric typewriter or a new word processing system. Of course, the system would have to be redesigned so that its energies would be channeled to secure the higher rate. This decision is the primary responsibility of a system's manager, as is any investment or re-configuration of the system.

Beer[2] suggests that differences of this type can be tracked by means of a ratio which he calls the "latency" ratio. For example, a ratio of a system's systemic output productivity to its bounded output productivity reveals how closely the system's existing capability approaches its potential with new investment. It is also a measure of the resource stewardship of the system's manager. A manager may raise process or bounded productivity by depleting and "squandering his latent resources." People and equipment may be overworked, "let go," under-maintained or reassigned in an effort to reduce costs. But the result will be a reduction of the capacity and capability of the system (i.e., a lower denominator in the bounded output productivity formula). This depletion results in a reduction in the ratio of systemic to bounded productivity and serves as a signal that short term improvements have been gained at the cost of losing future performance. That is, the ratio would reveal that the manager had *dis*-invested during the period.

SUMMARY

Productivity has been defined in systems terms. It was argued that depending on the assumptions made about the system's environment and management's perceived relation to it, three different concepts of productivity can be identified. Based on a theory of system's environments, notions of (1) process productivity, (2) bounded productivity, and (3) systemic productivity were developed and appropriate measures suggested. The result is a general systems theory of productivity which should aid the system's manager in measuring and improving the efficiency of any productive system.

REFERENCES

1. R. L. Ackoff and F. E. Emery, *On Purposeful Systems.* Aldine-Atherton, Chicago, 1972.
2. S. Beer, *Brain of the Firm.* Herder and Herder, New York, 1972.
3. C. W. Churchman, *The System's Approach.* Dell Publishing Co., New York, 1968.
4. C. W. Churchman, *The Design of Inquiring Systems.* Basic Books, New York, 1971.
5. J. R. Commons, *Institutional Economics.* The University of Wisconsin Press, Madison, Wisconsin, 1961, pp. 628–629.
6. C. E. Craig and R. C. Harris, "Total productivity measurement at the firm level." *Sloan Management Review*, **14,** No. 3, Spring 1973.
7. R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm.* Prentice Hall, Englewood Cliffs, New Jersey, 1963.
8. E. F. Denison, "The sources of economic growth in the United States and the alternatives before us." Committee for Economic Development, New York, 1962.
9. E. F. Denison, *Why Growth Rates Differ.* Brookings Institute, Washington, 1974.
10. E. F. Denison, *Accounting for United States' Economic Growth 1929–1969.* Brookings Institute, Washington, 1974.
11. S. Eilon, B. Gold and J. Soesan, *Applied Productivity Analysis for Industry.* Pergamon International Library, Oxford, 1976.
12. F. E. Emery and E. L. Trist, "The causal texture of organizational environments." *Human Relations*, **18,** 1963, pp. 20–26.
13. S. Fabricant, *A Primer on Productivity.* Random House, New York, 1969.
14. M. J. Farrell, "The measurement of productive efficiency." *Journal of the Royal Statistical Society*, **120,** Part II, 1957: pp. 253 290.
15. M. J. Farrell and M. Fieldhouse, "Estimating efficient productions under increasing returns to scale." *Journal of the Royal Statistical Society*, **125,** Part II, 1962, pp. 252 267.
16. I. Fisher, *The Nature of Capital and Income.* Macmillan, New York, 1906.
17. B. Gold, *Foundations of Productivity Analysis.* University of Pittsburgh Press, Pittsburgh, 1955.
18. B. Gold, *Explorations in Managerial Economics: Productivity, Costs, Technology and Growth.* Basic Books, New York, 1971.
19. L. Greenberg, *A Practical Guide to Productivity Measurement.* The Bureau of National Affairs, Washington, D.C., 1973.
20. A. P. Grosse, "The technological structure of the cotton textile industry." In *Studies in the Structure of the American Economy*, edited by W. Leontiof, *et al.*, Oxford University Press, New York, 1953, Chapter 10.
21. W. T. Jerome, *Executive Control—The Catalyst.* John Wiley, New York, 1961.
22. J. W. Kendrick, *Productivity Trends in the United States.* Princeton University Press, Princeton, 1961.
23. J. W. Kendrick and D. Creamer, *Measuring Company Productivity.* The Conference Board, New York, 1965.
24. C. H. Kriebel, *et al.*, "Modeling the productivity of information systems." Carnegie-Mellon University, Technical Report No. NSF APR 75-20546/76/TR2, Pittsburgh, April 1976 (Revised June 1976).
25. W. H. McWhinney, "Organizational form, decision modalities and the environment." *Human Relations*, **21,** No. 3, August 1968, pp. 269–281.
26. S. Melman, "The rise of Administrative overhead in the manufacturing industries of the United States, 1899 1947." *Oxford Economic Papers*, III, No. 1, January 1951.

27. S. Melman, *Dynamic Factors in Industrial Productivity.* Basil Blackwell, Oxford, 1956.
28. A. K. Rice, *Productivity and Social Organization.* Tavistock, London, 1970.
29. W. Seitz, "The measure of efficiency relative to a frontier production function." *American Journal of Agriculture Economics*, November 1970, pp. 505–511.
30. W. Seitz, "Productive efficiency in the steam-electric generating industry." *Journal of Political Economy*, July-August 1971, pp. 878–886.
31. A. Smith, *An Inquiry Into the Nature and Causes of the Wealth of Nations.* Great Books of the Western World, Vol. 39.

Dr. RICHARD O. MASON was born in Memphis, Tennessee. He received his B.S. from Oregon State University and his Ph.D. from University of California, Berkeley. He is currently Associate Professor of Management and Information Systems and Chairman, Public and Not-for-Profit Management at the Graduate School of Management, University of California, Los Angeles. Previously he served as Assistant Dean, Professional Masters Program at the Graduate School, U.C.L.A.

Dr. Mason served eight years as a Computer Specialist for Burroughs Corporation and two years, on leave from U.C.L.A., with the National Science Foundation in Washington, D.C.

Dr. Mason was the Editor of the Management and Behavioral Science Department for *Management Science* and is currently President-Elect of the Western Academy of Management.

Dr. Mason is a member of the Institute of Management Science, the Academy of Management, American Accounting Society and Association of Computing Machinery.

nt. J. General Systems
979, Vol. 5, pp. 31–43

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

SYSTEMS AND DISTINCTIONS; DUALITY AND COMPLEMENTARITY†

JOSEPH A. GOGUEN

Department of Computer Science, University of California, Los Angeles, California, U.S.A.

and

FRANCISCO J. VARELA‡

School of Medicine, University of Colorado, Denver, Colorado, U.S.A.

(*Received December 6, 1977; in final form May 26, 1978*)

This paper presents a framework within which a number of preferred system *descriptions* can be unified. Of particular interest are the differences stemming from the study of natural systems (biological and social) and man-made systems (engineering and computer). It is claimed that the insights from the study of natural systems have been developed less extensively and precisely. One important case discussed here is the relationship between describing a system as *autonomous* or as *controlled*. A precise characterization of *complementarity* in descriptions is proposed through the notion of adjoint functor, to explicate the interplay of these dual views. Other descriptive dualities of similar kind are discussed; the complementarity between nets and trees is presented in full detail.

INDEX TERMS Complementarity, distinctions, autonomy, control, adjoint functor, trees, networks.

. INTRODUCTION

The world does not present itself to us neatly divided into systems, subsystems, environments, and so on. These are divisions which we make ourselves, for various purposes, often subsumed under the general purpose evoked by saying "for convenience". It is evident that different people find it convenient to divide the world in different ways, and even one person will be interested in different systems at different times, for example, now a cell, with the rest of the world its environment, and later the postal system, or the economic system, or the atmospheric system.

The established scientific disciplines have, of course, developed different preferred ways of dividing the world into environment and system, in line with their different purposes, and have also developed different methodologies and terminologies consistent with their motivations.

In this paper, we present a framework within which a number of these various preferred views on systems can be unified. Of particular interest to us are the differences stemming from the study of natural systems (particularly biological and social systems) and man-made systems (such as engineering and computer systems). Contemporary system theory has developed extensively through experience in the later fields, but the insights derived from natural systems have remained by and large much less formally developed. We hold that the notions of cooperative interaction, self-organization, and autonomy—in brief, holistic notions—are relevant and basic to the study of natural systems. In the framework of this paper and its successors, these notions are not only made more explicit and applicable, but are also presented as complements to the more traditional notions of system theory, such as control and input-output behavioral description.

†This paper was partially supported by the Naropa nstitute, Boulder, Colorado (J.G.), and the Alfred P. Sloan oundation (F.V.). It is part of a series of three. The second aper will present some mathematical foundations for self-eferential processes in systems, including a Spencer–Brownian calculus of indicational re-entry. The third paper ill deal with applications to various specific cases.

‡Present address: Brain Research Laboratories, New York Jniversity Medical School, 550 First Avenue, New York City 0016.

Our interest and background for this work is derived from a variety of sources. Humberto Maturana and Heinz von Foerster have been instrumental in clarifying the importance of circular and cognitive processes in natural systems. Heinz von Foerster, Warren McCullough, Gordon Pask and Lotfi Zadeh have been influential in suggesting a broader vision of system theory and cybernetics. Saunders MacLane, Samuel Eilenberg and William Lawvere have provided crucial mathematical inputs. Moreover, the teachings of Chogyam Trungpa, Rinpoche have been a constant source of guidance and inspiration in understanding and putting together all these ideas. Finally, we acknowledge the positive influences of Gregory Bateson, Charlotte Linde, G. Spencer-Brown and the ADJ group.

The next section discusses in general terms the role which distinction plays in the creation and recognition of systems. The following three sections discuss certain dual perspectives on systems in some detail, including the autonomy/control, state–variable/input–output, holism/reductionism, and net/tree dualities. The fifth section develops the suggestion that such alternatives are complementary rather than antagonistic, into a suggestion that their inter-relationship can often be expressed precisely as an adjoint functor relationship, in the sense of categorical algebra. The final section discusses the holism/reduction relationship in some detail, in relationship to the philosophy of science.

Future papers will build on this foundation, to discuss the notion of autonomy in terms of a mathematical theory of self-reference or indefinite recursion and its applications to various systems.

2. DISTINCTION AND INDICATION

A *distinction* splits the world into two parts, "that" and "this", or "environment" and "system", or "us" and "them", etc. One of the most fundamental of all human activities is the making of distinctions. Certainly, it is the most fundamental act of system theory, the very act of defining the system presently of interest, of distinguishing it from its environment.

Distinctions coexist with purposes. A particularly basic case is a system defining its own boundaries and attempting to maintain them; this seems to correspond to what we think of as self-consciousness. It can be seen in individuals (ego or indentity maintenance) and in social units (clubs, subcultures, nations). In such cases, not only is there a distinction, but an *indication*, that is, a marking of one of the two distinguished states as being primary ("this", "I", "us", etc.); indeed, it is the very purpose of the distinction to create this indication.[1, 2]

A less basic kind of distinction is one made by a distinctor for some purpose of his own. This is what we generally see explicitly in science, for example, when a discipline "defines its field of interests", or a scientist defines a system which he will study.

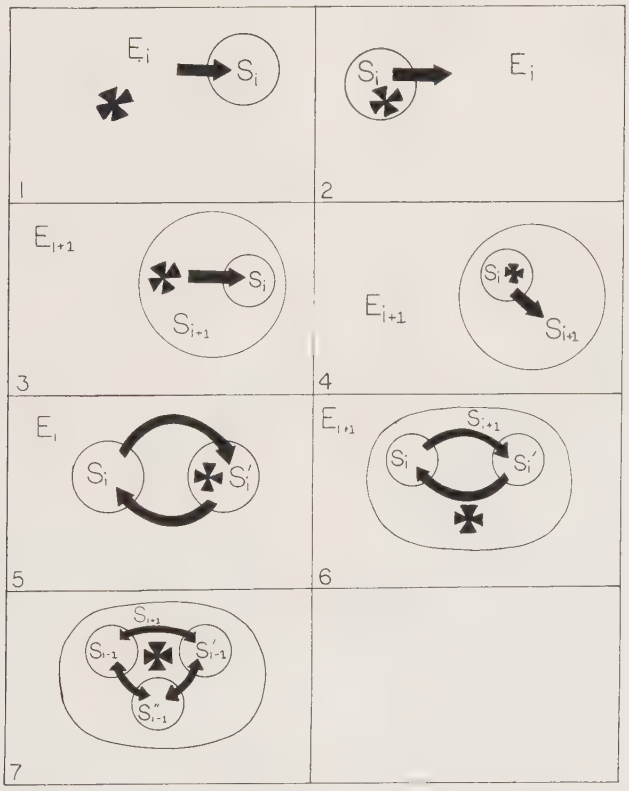
In either case, the establishment of system boundaries is inescapably associated with what we will call a *cognitive point of view*, that is, a particular set of presuppositions and attitudes, a perspective, or a frame in the sense of Bateson[3] or Goffman[4]; in particular, it is associated with some notion of value, or interest. It is also linked up with the cognitive capacities (sensory capabilities, knowledge background) of the distinctor. Conversely, the distinctions made reveal the cognitive capabilities of the distinctor. It is in this way that biological and social structures exhibit their coherence, and make us aware that they in fact have cognitive capacitities or that they are "conscious" in some degree.

The importance for system theory of cognitive coherence (or cognitive point of view, or cognitive capability) is a theme which runs throughout this paper and its successors. Because of the focus on system theory, we shall feel free to invoke the idea of an "observer", one or more persons who embodies the cognitive point of view which created the system in question, and from whose perspective it is subsequently described.

A simple but fundamental property of the situation involving a system and an observer, is that he may choose to focus his attention either on the internal constitution of the system, or else on its environment, taking the system's properties as given. That is, an observer can make a distinction into an indication through the imposition of his value. If the observer chooses to pay attention to the environment, he treats the system as a simple entity with given properties and seeks the regularities of its interaction with the environment, that is, the constraints on the behavior of the system imposed by its environment.[5] This leads naturally to the problem of controlling the behavior of the system, as considered in engineering control theory. On the other hand, the obser-

may choose to focus on the internal structure the system viewing the environment as baground, for example, as a source of perturions upon the system's autonomous behavior. om this viewpoint, the properties of the system erge from the interactions of its components. Biology has iterated this process of indication, ating a hierarchy of levels of biological study. e cell biologist emphasizes the cell's autonomy, d views the organism of which it is part as le more than a source of perturbations for ich the cell compensates. But the physiologist ws the cell as an element in a network of erdependencies constituting the individual orism: this corresponds to a wider view of vironment, namely the ecology in which the dividual participates. A population biologist kes his distinctions at a still higher level, and largely ignores the cell. A similar hierarchy of levels can be found in the social sciences. It seems to be a general reflection of the richness of natural systems that indication can be iterated to produce a hierarchy of levels.

At a given level of the hierarchy, a particular system can be seen as an outside to systems below it, and as an inside to systems above it; thus, the status (i.e., the mark of distinction) of a given system changes as one passes through its level, in either the upward or the downward direction. The choice of considering the level above or below corresponds to a choice of treating the given system as autonomous or controlled (constrained). Figures 1 to 7 illustrate a variety of configurations of systems, subsystems, and marks, and Figure 8 illustrates the hierarchy of levels.



FIGURES 1–7 Various configurations of systems, subsystems and marks; each configuration represents a cognitive viewpoint, and the mark indicates its center. The arrows indicate the flow of signals and interactions.

1—Control of a System S_i by its Environment E_i. 2—Autonomy of System S_i in its Environment E_i. 3—Control of a Subsystem S_i in a System S_{i+1}. 4—Autonomy of a Subsystem S_i of a System S_{i+1}. 5—Feedback Control of System S_i by System S_i'. 6—Communication between (Coordination of) Two Subsystems S_i, S_i' of System S_{i+1}. 7—Coordination (Ecology) of Subsystems S_{i-1}, S_{i-1}', S_{i-1}''.

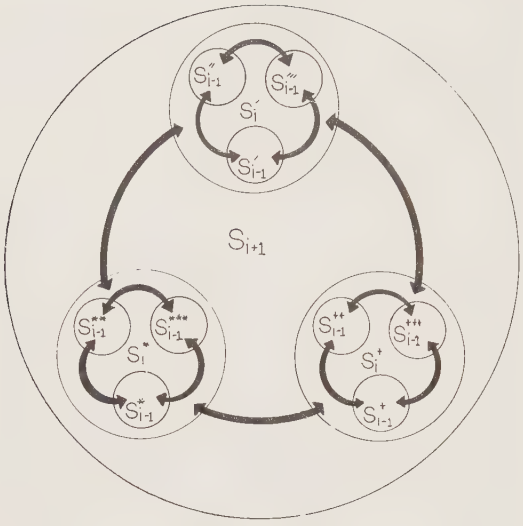


FIGURE 8 Diagramatic evocation of a hierarchy of system levels: Systems $S^*_{i-1}, S^{**}_{i-1}, S^{***}_{i-1}, \ldots$ of level $(i-1)$ constitute system S^*_i at level i; similarly, systems $S^*_i, S^+_i, S'_i, \ldots$ of level i constitute system S_{i+1} at level $(i+1)$; and S_{i+1} together with other systems of level $(i+1)$ will constitute a system at level $(i+2)$; and so on, upward and downward.

3. RECURSION AND BEHAVIOR

In system theory, the autonomy/control distinction appears more specifically as a recursion/behavior distinction. The behavioral view reduces a system to its input-output performance or behavior, and reduces the environment to inputs to the system. The effect of outputs on environment is not taken into account in this model of the system. The recursive view of a system emphasizes the mutual interconnectedness of its components.[6,2,7] That is, the behavioral view arises when emphasis is placed on the environment, and the recursive view arises when emphasis is placed on the system's internal structure.

If we stress the autonomy of a system S_i (see Figures 1 to 7) then the environmental influences become perturbations (rather than inputs) which are compensated for through the underlying recursive interdependence of the system's components (the S_{i-1}'s in the figures). Each such component, however, is treated behaviorally, in terms of some input-output description.

The recursive viewpoint is more sophisticated than the behavioral, since it involves the simultaneous consideration of three different level whereas the behavioral strictly speaking involve only two. This is because the behavioral mode in taking the environment's view of the systen does not involve making any new distinction But expressing interest in how the syste achieves its behavior through the interdepende action of its parts adds a new distinction, b tween the system and its parts.

The following may help to make this seem le abstract. The most traditional way to express th interdependence of variables in a system is b differential equations. A (time varying) aut nomous system can be represented by equatio of the form

$$\dot{x}_i = F_i(x,t), \quad \text{for } 1 \leqq i \leqq n \qquad (1$$

where $x = \langle x_1, \ldots, x_n \rangle$ is the state vector of th system. The autonomous behavior of the syste is described by a solution vector $x(t)$, whic satisfies (1). This involves treating everything happening on the same level, and all variables being observable; in effect, environment is treate as part of the system (or ignored).

However, the effect of the environment on t system can be represented by a vector $e = \langle e_1, \ldots e_k \rangle$ of parameters, giving

$$\dot{x}_i = F_i(x, e, t), \quad \text{for } 1 \leqq i \leqq n,$$

which explicitly takes account of two leve Solutions to the system (2) are now also par meterized by e, that is, they are of the for $x(e,t)$.

The situation of (2) can be elaborated in tw directions. In control theory, it is usual to assun that the internal variables x of the system a either unobservable, or are of no direct interes and that we have instead direct access to (interest in) an output vector y of variables whic are functionally dependent on x. The variables are usually taken to be under the control of th observer, and the question is posed, how to us those variables to obtain certain desired values y. The equations are thus of the form

$$\dot{x}_i = F_i(x, e, t)$$
$$y = H(x). \qquad (3$$

Strictly speaking, the equations span three leve and can be used, for example, to infer info

ation about the system's internal state, but the mphasis ("mark") is on the environment, which identified with the observer. Behavior appears an input–output function $y(e,t)$, the observle results of applying the inputs (also called ontrols") e to the system.

An alternative elaboration of the situation of views the vector e as not necessarily or rticularly under the control of an observer, but ther as a source of perturbations upon (2). For ample, the components e_j of e may be some efficients which are regarded as constants in e original equation (1). A natural question to se is the stability of the system under such rturbations, that is the relation of (1) to a rturbed system

$$\dot{x}_i = F_i(x,e,t) + \delta F_i(x,e,t) \tag{4}$$

which δ (in a fairly intuitive notation) repreits a "small change". It is known, for example, at changes in structural constants can cause the stem to undergo a "catastrophic" change (in the nse of Thom[8]) into a new configuration.

The system (2) has in it nothing which intrinsilly prefers the approach of either (3) or (4). his choice depends on the interest of the alyst.

Note that "recursion" (in the broad sense of edback which affects the future) plays a role in these formulations, but is more obscure in the ntrol theory interpretation. On the other hand, e behavioral information, though still available, more obscure in the stability interpretation (4). e are not, of course, claiming that either of ese approaches is inherently better.

Historically speaking, some of the many posle approaches to systems have been much ore developed than others. The most highly veloped parts, in fact, center on the notions of ntrol, input–output behavior, and statensition. This is presumable because of the erest in applying these approaches in gineering.

It seems, however, that the notion of autonomy particularly important for natural systems, e.g., ological and social systems, and the lack of a ll-developed theory of autonomous systems is t as a serious difficulty. An engineer designing artifact will choose the inputs of interest to n for this application with some assurance that e choice will be adequate. But a biologist udying a cell is forced to acknowledge the autonomy of the cell; if the biologist's preferences for input and output variables do not match the cell's internal organization, the biologist's theory will not work. Furthermore, the hierarchy of levels seems to particularly assert its importance for natural systems, so that it is generally necessary to take account of at least three levels. Even when the lowest level is very well understood, the role which it plays, at the next higher level when interconnected with other systems, can be quite obscure. For example, an enzyme biochemist may be able to describe a particular metabolic loop very effectively by a transfer function, but be quite unable to specify how it fits into the overall metabolic process of the cell as a coherent whole.

This situation of being unable to understand how elements, even quite well understood elements, coordinate or somehow function effectively together at the next higher level, is quite common in the study of natural systems, and is a major source of our desire for a better developed theory of autonomous systems.

Some fragments of theory emphasizing the autonomy of systems do exist, but they are far less developed than is the behavioristic approach. First and foremost, the idea of stability derived from classical mechanics, has been extensively studied and used. As we said before, a set of interdependent differential equations can be used to represent the autonomous properties of a whole system. Rosen,[9] Iberall[10] and Lange[11] have applied this perspective to natural systems with various degrees of emphasis on autonomous behavior. More specific examples can be found in population biology,[12] cellular biology,[13,14] and more recently, in neurobiology.[15] Some thought has been given to cooperative interactions in this area of hierarchical multilevel systems. The idea of hierarchy is often presented from the point of view of the interdependency of different levels of systems descriptions.[16,17,18] Particular instances of hierarchical structure including multilevel cooperation can be found in [19,20,21]. Goguen[22,23] presents a general theory of hierarchical systems of interdependent processes. Its basic ideas are interconnection, behavior, and level, and its theoretical framework is categorical algebra. A last area in which the idea of a whole system is somewhat explicit is that of self-organizing systems. Work in this area based on an information-theoretic approach includes von Foerster,[24] and Atlan.[25]

We do not intend to unite all these various threads of research together in a single framework. Rather, we emphasize the ways in which pairs of seemingly different points of view, such as autonomy/control, are complementary, in the sense of contributing to a better understanding of natural systems. But the idea of complementarity, fundamental though it seems, is still vague. The next two sections develop an explicit definition.

4. NETS AND TREES

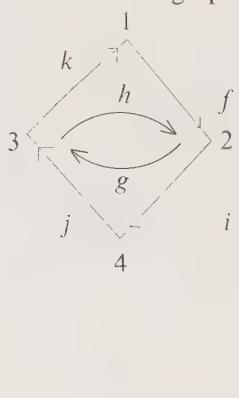
If we retain interest only in the connectivity of a system, it is possible to represent the duality recursion/behavior by a network/tree duality. Intuitively, the nodes in these nets or trees represent the elements or components of a system, while their links represent interactions or interconnections. The reciprocal connectivity of a net suggests the coordination of a system's elements; a tree structure suggests the sequential subordination of a system's parts, each part having its own well-defined input–output behavior description. To be sure, in retaining only the basic connectivity of a system's organization, much is discarded in the net/tree representation. We intend to use this convenient general representation to study complementarity.

Now to the definition of nets and trees. Let there be a set of $\{v_1,\ldots,v_n\}$ of nodes (components or parts), which are to be interconnected by a set of $E=\{e_1,\ldots,e_p\}$ edges (relations or processes).

1. DEFINITION A *network* is a directed graph G, that is, a quadruple $G=\langle|G|, E, \partial_0, \partial_1\rangle$, where $|G|=\{v_1,\ldots,v_n\}$, and $\partial_i: E\to|G|$ are the *source* ($i=0$) and *target* ($i=1$) functions, from the edges to the nodes of G. If $e\in E, \partial_0 e=v$ and $\partial_1 e=v'$, then we write $e: v\to v'$. ■

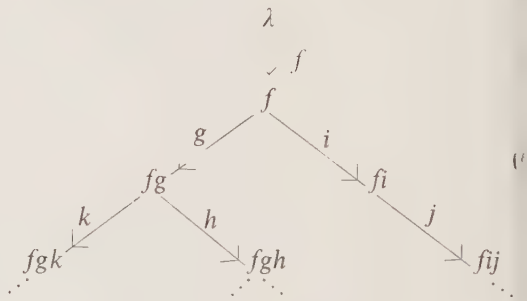
2. DEFINITION A *path* from v to v' in a graph G is a finite sequence $p=e_0\ldots e_n$ of edges which are adjacent, that is, satisfy $\partial_1 e_i=\partial_0 e_i$, for $1\leqq i\leqq n$, with $\partial_0 e_0=v$ and $\partial_1 e_n=v'$. If $\partial_0 p=v$ and $\partial_1 p=v'$, then write $p: v\to v'$. ■

3. EXAMPLE Consider the graph G:



(5)

The nodes 1, 2, 3 might represent three physica locations, each with a CB (Citizens Band) radio Because of differences in transmitter and receive strength, available frequencies, and terrain, com munication is possible only along the channel indicated by arrows in (5). For example, there i a mountain between 1 and 4. Let us assume tha node 1 is of particular interest –say it is ou base. Then we are interested in the patterns o transmission which are possible starting from ou base. For example, to reach node 3, if channel is out, we can send a message via fij; to verify it correctness, it could be sent back to node 1, vi the path $fijk$; node 3 might also generate a reply which would require a message to be sent t node 2, giving a path $ijkf$; and so on. Thus, w are interested in the set of all paths in G wit source 1. This collection itself has a branchin structure, because starting with a given path, i can sometimes be developed further by choosin alternative edges to get alternative paths. Th collection of all choices can be represented by th following tree



Notice that it is an infinite tree ■

In some sense the tree (6) "unravels" or "u folds" the graph (5) from node 1. To make th more precise we define trees, pointed graphs, an structure-preserving mappings of graphs calle graph homomorphism.

4. DEFINITION A *pointed graph* G is a 5-tup $\langle|G|, E, \partial_0, \partial_1, a\rangle$ such that $\langle|G|, E, \partial_0, \partial_1\rangle$ is graph and $a\in|G|$ is a vertex.

A pointed graph is *reachable* if for each verte $v\in|G|$ there is a path $a\to v$ in G.

A graph G is *loop-free* if for all $v, v'\in|G|$ the is at most one path $v\to v'$.

A *tree* is a reachable loop-free pointed graph ■

5. DEFINITION Let G be a pointed grap $\langle|G|, E, \partial_0, \partial_1, a\rangle$. Then the *unfoldment* $U_a(G)$ of

from a, is the graph in which: $|U_a(G)|$ is all the paths $p: a \to v$, for $v \in |G|$; the edges of $U_a(G)$ are the pairs $\langle p, pe \rangle$, such that $p, pe \in |U_a(G)|$, and $e \in E$; $\partial_0 \langle p, pe \rangle = p$, and $\partial_1 \langle p, pe \rangle = pe$. The null path $a \to a$ is written $1_a: a \to a$, and is taken to be the point for $U_a(G)$ ■

6. PROPOSITION *Let G be a pointed graph $\langle |G|, E, \partial_0, \partial_1, a \rangle$, and $U_a(G)$ the unfoldment of G from a. Then $U_a(G)$ is a tree.*

Proof We must show that $U_a(G)$ is a pointed, reachable, and loop-free graph. By the definition of $U_a(G)$, it has point $1_a: a \to a$.

We now show that $U_a(G)$ is reachable. Consider a node $p: a \to v$ of $U_a(G)$, say $p = e_0 \dots e_n \neq 1_a$, with $e_i \in E$. Then we can show that:

$$q = \langle 1_a, e_0 \rangle \langle e_0, e_0 e_1 \rangle \langle e_0 e_1, e_0 e_1 e_2 \rangle \dots \langle e_0 \dots e_{n-1}, e_0 \dots e_n \rangle$$

is a path from 1_a to p in $U_a(G)$. Clearly its source is 1_a since

$$\partial_0 q = \partial_0 \langle 1_a, e_0 \rangle = 1_a,$$

and its target is p since

$$\partial_1 q = \partial_1 \langle e_0 \dots e_{n-1}, e_0 \dots e_n \rangle = e_0 \dots e_n = p.$$

Moreover q is a path since its edges are adjacent, that is,

$$\begin{aligned} &\partial_1 \langle e_0 \dots e_k, e_0 \dots e_{k+1} \rangle \\ &= \partial_0 \langle e_0 \dots e_{k+1}, e_0 \dots e_{k+2} \rangle \\ &= e_0 \dots e_{k+1}, \text{ for } 0 \leqq k \leqq n-1. \end{aligned}$$

So there is a path from 1_a to every node in $U_a(G)$.

Last we show that $U_a(G)$ is loop-free, that is, for every pair of nodes p, p' in $U_a(G)$ there is at most one path $p' \to p$ in $U_a(G)$. Consider again $p = e_0 \dots e_n$, and lets show first that there is exactly one edge with target p, e.g. exactly one path of length one. Any edge with target p is of the form $\langle r, re_n \rangle$, with $re_n = p$. Thus r must equal $e_0 \dots e_{n-1}$, and the unique edge is $\langle e_0 \dots e_{n-1}, p \rangle$. This says that if $p' \neq p$, a path $p' \to p$ must end with edge $\langle e_0 \dots e_{n-1}, p \rangle$. Let now $p_k = e_0 \dots e_k$. Then a path $p' \to p$ must be a composite of a path $p' \to p_{n-1}$ with edge $\langle p_{n-1}, p \rangle$. But we may reason in a similar way for the node p_{n-1}, and the path $p' \to p_{n-1}$, p_{n-2}, and $p' \to p_{n-2}$, and so on. Eventually we must find that $p' = p_k$ for some k, and the unique path $p' \to p$ is of the form

$$\langle p_k, p_{k+1} \rangle \langle p_{k+1}, p_{k+2} \rangle \dots \langle p_{n-1}, p \rangle.$$

If $p = p'$, the unique path $p \to p'$ is the null path at p. Thus $U_a(G)$ is loop-free, and the proof is complete ■

7. DEFINITION Let $G = \langle |G|, E, \partial_0, \partial_1 \rangle$ and $G' = \langle |G'|, E', \partial'_0, \partial'_1 \rangle$ be graphs. Then a *graph morphism* is a pair $\langle |F|, F \rangle$ of functions $|F|: |G| \to |G'|$ and $F: E \to E'$, such that the source and target relationships are preserved, that is, such that

$$\partial'_0(F(e)) = |F| \partial_0(e),$$

and

$$\partial'_1(F(e)) = |F| \partial_1(e);$$

i.e. such that the diagram

$$\begin{array}{ccc} & F & \\ E & \longrightarrow & E' \\ \partial_i \downarrow & & \downarrow \partial_{i'} \\ & |F| & \\ |G| & \longrightarrow & |G'| \end{array}$$

commutes for $i = 0$ and $i = 1$. We will generally abbreviate $\langle |F|, F \rangle$ to just F. A *morphism of pointed graphs*, from G to G', is a graph morphism F such that $|F| a = a'$, where a, a' are the selected vertices of G, G' respectively ■

The relation between a graph G and its unfoldment is, from our perspective, very interesting. Given a node a in G, then $U_a(G)$ is a loop-free version of G. We could say $U_a(G)$ expresses G as a (possibly infinite) chain of subordinated choices, starting from the selected node. The unfoldment of G optimally "covers" G in a sense which is made precise through the "universal property" of $U_a(G)$, that any graph morphism $F: T \to G$ can be factored through a "covering morphism" $C_G: U_a(G) \to G$, defined as follows: for p a node of $U_a(G)$, let $|C_G|\ p = \partial_1 p$; and for $\langle p, pe \rangle$ an edge of $U_a(G)$, let $C_G \langle p, pe \rangle = e$.

We now show that C_G is a graph morphism.

For $\langle p, pe\rangle$ an edge of $U_a(G)$, then

$$\partial_0 C_G\langle p, pe\rangle = \partial_0 e,$$

$$|C_G|\partial_0\langle p, pe\rangle = |C_G|p = \partial_1 p,$$

and

$$\partial_0 e = \partial_1 p$$

because pe is a path. Also,

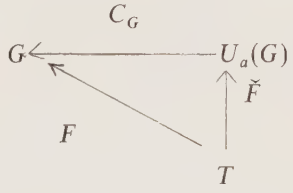
$$\partial_1 C_G\langle p, pe\rangle = \partial_1 e,$$

and

$$|C_G|\partial_1\langle p, pe\rangle = |C_G|\ pe = \partial_1 pe = \partial_1 e.$$

We now show that any other morphism from a tree can be factored through C_G.

8. THEOREM *Let G be a pointed graph, let T be a tree, and let $F: T \to G$ be a pointed graph morphism. Then there is a unique pointed graph morphism $\check{F}: T \to U_a(G)$ such that:*



commutes.

Sketch of Proof Let $v \in |T|$. Then there is a unique path $p: r \to v$ in T, where r is its root, and $Fp: Fr = a \to Fv$ is a path in G. Now we define $|\check{F}|v = Fp$, and for $e: v \to v'$, an edge in T, we define $\check{F}e = \langle Fp,\ (Fp)Fe\rangle$, and edge in $U_a(G)$. Then:

$$|C_G||F|v = |C_G|Fp = F\partial_1 p = |F|v,$$

and

$$C_G\check{F}e = C_G\langle Fp, (Fp)Fe\rangle = Fe.$$

Thus

$$C\check{F} = F.$$

We now verify that $\check{F}$ is a pointed graph morphism:

$$\partial_0\check{F}e = \partial_0\langle Fp, (Fp)Fe\rangle = Fp, |\check{F}|\partial_0 e = |\check{F}|v = Fp,$$

and

$$\check{F}r = F1_r = 1_a;$$

$$\partial_1\check{F}e = \partial_1\langle Fp, (Fp)Fe\rangle = (Fp)Fe = F(pe) = |\check{F}|\partial_1 e.$$

Thus $\check{F}$ satisfies the required conditions.

For a detailed proof, and the uniqueness argument see [26] Theorem 7 ■

This theorem brings into focus the basic intuition that there is a mutual interdependence between a system's elements (as a graph), and the sequential subordination of their interconnections (as a tree). To express this more clearly, let $\mathbf{Gr}_*$ be the class[27] of pointed graphs, and let $\mathbf{T}$ be the class of trees. Also, if G and G' are pointed graphs, let $\mathbf{Gr}_*(G, G')$ denote the set of all pointed graph morphisms from G to G'.

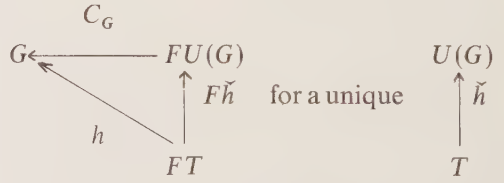
Since every tree is by definition a pointed graph, we have a mapping

$$\mathbf{T} \xrightarrow{\ F\ } \mathbf{Gr}_*$$

which simply views trees as pointed graphs. We also have a mapping

$$\mathbf{Gr}_* \xrightarrow{\ U\ } \mathbf{T}$$

which assigns to every pointed graph the tree $U(G)$ which covers it. These two mappings F, U are tightly interlocked. For any $h \in Mor\ (G, FT)$ we have



by Theorem 8. This says that there is a bijection

$$\varphi: \mathbf{Gr}_*(FT, G) \to \mathbf{Gr}_*(T, UG)$$

defined by $\varphi(h) = \check{h}$. We will call (F, U, φ) a *complementarity* between $\mathbf{Gr}_*$ and $\mathbf{T}$.

This notion of net/tree complementarity effectively relates two levels of description of systems, in such a way that each necessitates the other. It is convenient at this point to see that similar notions of complementarity apply to other situations, and so we now turn to the general notion.

5. COMPLEMENTARITY AND ADJOINTNESS

The net/tree complementary is a particularly clear instance of the interdependence of apparent dualities. This Section develops this idea in the

neral setting of category theory, which is be-
ming increasingly useful in systems theory.[28,29]
eaders unfamiliar with this terminology may
d a leisurely introduction in [30,31] or [32]: we
empt to stay at a fairly intuitive level in this
ction, although some technicalities are
evitable.

The intuitive idea of a *category* is that it
nbodies some structure by exhibiting the class
all *objects* having that structure, together with
the structure preserving mappings or *mor-
isms* among them. (Somewhat more technically,
tegories assume there is an associative oper-
on of composition on those morphisms whose
urce and target match.) This idea is due to
lenberg and MacLane.[33]

For example, pointed graphs and pointed
aph morphisms constitute a category. If **C** is a
tegory, and A, B are objects in **C**, we shall let
$A, B)$ denote the set of all morphisms in **C**
om A to B.

Usually, we are interested not only in objects
om various categories, but we are even more
terested in certain constructions performed on
e objects of one category to yield objects of
other category. For example, unfoldment is a
nstruction performed upon graphs which yields
es. This construction has a kind of consistency,
that it can also be extended to the morphisms;
at is, a morphism of pointed graphs induces in
natural way, a morphism between their unfold-
nts. This kind of consistency is expressed by
ying that the construction is *functorial*, or is a
nctor. (More technically, this has to do with the
eservation of the composition of morphisms).

However, the unfoldment construction is na-
ral in a much stronger sense: the "optimal"
vering of a graph is its unfoldment; this is
pressed by the universal property of
heorem 8, and the bijection φ of the previous
ction. The concept of *adjunction* generalizes just
is state of affairs, to the following situation:

Let **A** and **B** be categories, let F be a functor
om **A** to **B**, and G a functor from **B** to **A**. Then
adjunction is, in addition, a natural bijection.

$$\varphi : \mathbf{B}(FA, B) \to \mathbf{A}(A, GB).$$

This says that every morphism $f\colon FA \to B$ de-
rmines a unique morphism $\varphi(f)\colon A \to GB$. (The
ecise sense of the "naturalness" of φ is that of
tural transformation, due to Eilenberg and
acLane,[33] which however, we shall not define
re. The idea of adjunction is due to Kan.[34])

The discussion at the end of the previous section shows that the net/tree complementarity is an instance of the concept of adjunction. *What we now propose, is to explore the view that the precise concept of adjunction is an explication of the general (and vague) concept of complementarity.*

Another example of this is Goguen's adjunction between minimal realization and behavior.[28,35] Let **A** be the category of automata (in some fixed sense which we shall not explain in detail), and let **B** be the category of input–output behaviors of such automata (with appropriate morphisms). Then there is a functor from **B** to **A** which constructs the minimal automation $M(B)$ having the behavior B; and there is a functor Be from **A** to **B** which constructs the behavior $Be(A)$ of an automaton A. Moreover, there is a natural bijection

$$\varphi : \mathbf{B}(Be(A), B) \to \mathbf{A}(A, M(B))$$

which expresses the complementarity of the notions of internal state transition (as embodied in automata) and input–output behavior. Goguen[35] has shown that many other classes of systems exhibit such a complementarity with their input–output behaviors.

Here is still another example. If G is a graph, the collection of all paths (from all sources) in G forms a category whose objects are the nodes of G, and whose morphisms are the paths of G; this category is denoted $\mathbf{Pa}(G)$, and called the *path category* of G. **Pa** is a functor from the category **Gr** of graphs to the category **Cat** of (small) categories. There is also a functor F from **Cat** to **Gr**, which merely forgets the additional structure which categories have over graphs (namely, the possibility of composing morphisms), regarding the objects as nodes, and the morphisms as edges. Again, there is a natural bijection

$$\varphi : \mathbf{Cat}(\mathbf{Pa}(G), \mathbf{C}) \to \mathbf{Gr}(G, F\mathbf{C})$$

expressing the complementarity of graphs and categories. Alternatively, $\mathbf{Pa}(G)$ is the free category generated by the graph G, and the adjunction (or the corresponding "universal property") expresses this relationship. (See[26] for further details.)

In general, it is more satisfactory to consider different modes of description that may appear as opposites, to be complementary instead. This is

the case, quite rigorously, with the apparent dualities net/tree and recursion/behavior, as we have seen above. On a more intuitive level, there is a similar relationship for the pair autonomy/control discussed in earlier sections.

Lawvere[36] is a particularly fundamental, stimulating paper, which suggests that there is a complementary relationship between the traditional conceptual/formal viewpoints in the foundations of mathematics. This duality also appears as a semantics/syntax pair, in that Lawvere[37] has shown an adjunction between a functor which associates to each algebraic theory its category of semantic models (i.e., its algebras), and a functor which extracts from each category, the optimal syntactic theory of its algebraic component of structure.

Varela[38] has described many situations where apparent opposites are in fact complementary. Our usual way of thinking about pairs of descriptive notions is biased in the direction of regarding them as mutual negations, of the general form: not—A/A. It is the case, however, that many of the most important such pairs operate by mutual specification, that is, are complementary. We mention in particular, content/form, semantics/syntax, autonomy/control, net/tree, minimal-model/behavior, simultaneous/sequential, and environment/system. Of course, the germinal example is the female/male complementarity, which seems to be particularly misunderstood as an opposition in our contemporary culture.

This general point seems particularly clear in the context of systems theory: there is no whole system without an interconnection of its parts; and there is no whole system without an environment. Such pairs are mutually interdependent: each defines the other. What is remarkable about the notion of adjoint functor, is that it captures the notion of complementarity in a very precise way, without imposing any particular model for the nature of the objects so related. It is also worth noting that there is a well-developed theory of adjunctions; for example, the composition of two adjoint pairs of functors is another adjoint pair. Such results should be very useful in system theory. Of course, not all pairs of descriptive modes are complementary, and similarly, not all pairs of functors are adjoint. The so-called "adjoint functor theorem" provides some general conditions for when a given functor in fact does have an adjoint, and again, this may well find some application in general discussion about system theory. Much more work, including many further examples, will be needed to discover the proper domain of application, and the limits, of the adjointless idea.

The general system theory of Goguen[22,23] involves a hierarchy of levels, much as pictured in Figure 8, with functors going outward[39] which regard a component at a lower level as a whole system at the next higher level, and functors going inward which compute the behavior of the whole system, viewing the result as a single object at the lower level. There is a base level of given "objects" out of which systems can be constructed, and objects at level $(i+1)$ are interconnections (that is systems) of objects at level i. Goguen shows that each pair of outward/inward functors is an adjunction. The inward functor is in fact the fundamental categorical construction known as "limit". Goguen also shows that the construction of interconnecting a system of systems (over some common subparts as "terminals") to get a single system, is given by the dual concept of "colimit", which also appears as an adjunction. This is not the place to give details, but the connection with the themes of this paper should be evident.

6. HOLISM AND REDUCTIONISM

If we think of the philosophy of science, the duality holism/reductionism comes to mind as analogous to the material previously discussed in this paper.

Most discussions place holism/reductionism in polar opposition.[40,41,42] This seems to stem from the historical split between empirical sciences viewed as mainly reductionist or analytic, and the (European) schools of philosophy and social science that grope toward a dynamics of totalities.[43,44]

In the light of the previous discussion, both attitudes are possible for a given descriptive level and in fact they are complementary. On the one hand, one can move down a level and study the properties of the components, disregarding their mutual interconnection as a system. On the other hand, one can disregard the detailed structure of the components, treating their behavior only as contributing to that of a larger unit. It seems that both these directions of analysis always coexist either implicitly or explicitly, because these de

scriptive levels are mutually interdependent for the observer. We cannot conceive of components if there is no system from which they are abstracted; and there cannot be a whole unless there are constitutive elements.

It is interesting to consider whether one can have a measure for the degree of wholeness of a system. One can, of course, always draw a distinction, make a mark, and get a "system" but the result does not always seem to be equally a "whole system", a "natural entity", or a "coherent object" or "concept". What is it that makes some systems more coherent, more natural, more whole, than others? (See[45,7,46], for different ways to answer these questions.)

One thing to notice is, that in the hierarchy of levels, "emergent" or "immanent" properties appear at some levels. For example, let us consider music as a system or organization of notes (for the purpose of this example, we do not attempt to reduce notes to any lower level distinctions). Then harmony only arises when we consider the simultaneous or parallel sounding of notes, and melody only arises when we consider the sequential sounding of notes. That is, harmony and melody are emergent properties of a level of organization above that of the notes themselves. Similarly, form can only emerge at a still higher level of organization, relating different melodic units to one another. These properties, form, melody and harmony, are *systems properties*, arising from hierarchical organizations of notes into pieces of music; they are not properties of notes. (See[47] for further discussion of the hierarchical organization of music.) It also appears that "life" is an emergent property of the biological hierarchy of levels: it is nowhere to be found at the level of atoms and molecules; but it becomes clear at the level of cells through the autopoietic organization of molecules.[48,49] Language can be seen as an emergent property at a still higher level of this hierarchy.[50]

Thus, one point of view toward wholeness is that it co-occurs with interesting *emergent* properties at some level. A sequence of notes is whole if it is an interesting melody. Similarly, the significant differences in behavior of cells and organism from the behaviors of atoms and molecules, marks the cells and organisms as whole systems.

Another point of view toward wholeness, is that it can be measured by the *difficulty of reduction*: Because it is very hard to reduce the behavior of organisms to the behaviors of molecules, we may say that organisms are whole systems. Similarly, it is very difficult (if not impossible) to reduce the effects of melodies to the effects of notes. One must consider properties of patterns of notes or molecules.

A third point of view is that a system is whole to the extent that its parts are *tightly interconnected*, that is, to the degree that it is difficult to find relatively independent subsystems. This is clearly related to the previous views. An interesting corollary of this view is that a system with a strongly hierarchical organization will be less whole than a system with a strongly heterarchical organization; that is, nets are more whole than trees. More precisely, given that the graph of connections of the parts of a system has no isolated subsystems, the more tree-like it is, the less whole it is, while still being (presumably) a system. The extreme is probably a pure linear structure, without any branching at all.

A fourth point of view, is that a system seems more whole if it is more *complex*, that is, more difficult to reduce to descriptions as interconnections of lower level components. It is necessary in this discussion to take account of the very modern point of view that the relatively more complex a system is to describe, the more *random* it is; in fact, Kolmogorov,[51] who gave the original foundation for probability theory in terms of measure theory,[52] has proposed to redefine probability theory and information theory in terms of complexity theory. Thus, for example, the wholeness of a living system is, in everyday encounters, construed as unpredictability. The more difficult it is to reduce a system to a simple input/output control the more likely it is we will deem it alive. In this sense complete autonomy is logically equivalent to complete randomness. Another example: a piece of music which is too complex (relative to our cultural expectations and inherent capacities) will sound random, chaotic, perhaps meaningless, but it will also sound whole. Here the extreme is white noise. (See[47] for a discussion of hierarchical complexity with applications to musical aesthetics.) This viewpoint toward wholeness involves measurement relative to some standard interpreting system, such as a human being. But given such a standard, this viewpoint can be deduced from the preceding viewpoints. For surely, if it is difficult to describe a system, it will also be difficult to reduce it to lower levels, and its parts will seem to be tightly

interconnected. Quite possibly, its very complexity will appear as an emergent property.

On the other hand, a different cognitive viewpoint might well be better able to process what now seems like a very complex system, and thus see it as less whole. Once again, the relativity to cognitive capacity appears. This can be seen over and over in the history of science, as new techniques, tools, or ideas, suddenly make long standing problems soluble, or perhaps, uninteresting or even ill-posed.

These descriptive levels haven't been generally realized as complementary largely because there is a difference between publicly announced methodology and actual practice, in most fields of research in modern science. A reductionist attitude is strongly promoted, yet the analysis of a system cannot begin without acknowledging a degree of coherence in the system to be investigated; the analyst has to have an intuition that he is actually dealing with a coherent phenomenon. Although science has publicly taken a reductionist attitude, in *practice* both approaches have always been active. It is not that one has to have a holistic view as opposed to a reductionist view, or *vice versa*, but rather that the two views of systems are complementary.

There is a strong current in contemporary culture advocating "holistic" views as some sort of cure-all. This is easily seen in discussions of environmental phenomena, health delivery, and value systems. The way in which holism is used in this paper is quite different. We simply see reductionism and holism as complementary, or "cognitively adjoint", for the descriptions of those systems in which we are interested. Reductionism implies attention to a lower level, while holism implies attention to a higher level. These are intertwined in any satisfactory description; and each entails some loss relative to our cognitive preferences, as well as some gain.

REFERENCES AND NOTES

1. G. Spencer-Brown, *Laws of Form.* Bantam Books, New York, 1973.
2. F. Varela, "A Calculus for Self-Reference." *Int. J. Gen. Systems*, **2,** 1975, pp. 5–24.
3. G. Bateson, "A Theory of Play and Fantasy." Reprinted in: G. Bateson, *Steps Towards an Ecology of Mind.* Ballantine, New York, 1972.
4. E. Goffman, *Frame Analysis.* Harper, Colophon Books, New York, 1974.
5. Calling S "the system" rather than "the environment" already indicates a preference for marking S; that is, the language incorporates the preference. But we may speak of "marking the environment" to suggest that there are in fact two distinct possibilities.
6. H. von Foerster, "Notes for an Epistemology of Living Things." *L'Unite de L'homme*, edited by E. Morin and M. Piatelli, Eds. du Seuil, Paris, 1974, pp. 401–417.
7. F. Varela and J. Goguen, "The Arithmetic of Closure." *J. Cybernetics*, **8,** 1978.
8. R. Thom, *Stabilité Structurelle et Morphogénèse.* Benjamin, New York, 1972.
9. R. Rosen, *Dynamical Systems Theory in Biology.* John Willey, New York, 1972.
10. A. Iberall, *Towards a General Science of Viable Systems.* McGraw Hill, New York, 1972.
11. O. Lange, *Wholes and Parts.* Pergamon Press, New York, 1965.
12. R. May, *Model Ecosystems.* Princeton U. Press, Princeton, 1971.
13. M. Eigen, "Self-organization and the Evolution of Matter." *Naturwiss*, **58,** 1971, pp. 465 523.
14. B. C. Goodwin, *Temporal Organization of Cells.* Academic Press, New York, 1968.
15. A. Katchalsky, V. Rowland and R. Blumenthal, "Dynamic Patterns of Brain Cell Assemblies." *Neurosciences Res. Prog. Bull.*, **12,** No. 1, 1974.
16. H. Pattee, ed., *Hierarchy Theory.* George Brazillier, New York, 1972.
17. L. Whyte, A. Wilson and D. Wilson, *Hierarchical Structures.* Elsevier, New York, 1968.
18. M. Mesarovic, D. Macko, Y. Takahara, *Theory of Hierarchical Multilevel Systems.* Academic Press, New York, 1972.
19. S. Beer, *The Brain of the Firm.* Allen Lane, London, 1972.
20. L. Kohout and B. Gaines, "Protection as a General Systems Problem." *Int. J. Gen. Systems*, **3,** 1976, pp. 3–24.
21. T. Baumgartner, T. Burns, L. Meeker and B. Wild, "Open Systems and Multi-Level Processes: Implications for Social Research." *Int. J. Gen. Systems*, **3,** 1976, pp. 25–42.
22. J. A. Goguen, "Mathematical Representation of Hierarchically Organized Systems." *Global Systems Dynamics*, edited by E. Attinger, S. Karger, Basel, 1971, pp. 112–128.
23. J. A. Goguen, "Systems and Minimal Realization." *Proceedings of 1971 IEEE Conference on Decision and Control*, Miami Beach, 1972, pp. 42–46.
24. H. von Foerster, "On Self-Organizing Systems and Their Environments." *Self-Organizing Systems*, edited by M. Yovits and S. Cameron, Pergamon Press, London, 1966, pp. 31–50.
25. H. Atlan, *L'Organization Biologique et la Theorie de la Information.* Herman, Paris, 1972.
26. J. A. Goguen, "On Homomorphisms, Correctness, Termination, Unfoldments, and Equivalence of Flow Diagram Programs." *Journal of Computer and System Sciences*, **1,** No. 3, 1974, pp. 333–365.
27. Such a "class" is a sort of higher logical type, consisting on all sets satisfying some conditions. Such a notion has been formalized within Gödel–Bernays set theory, and does not lead to contradictions.
28. J. A. Goguen, "Realization is Universal." *Mathematical System Theory*, **6,** No. 4, 1973, pp. 359–374.
29. M. Arbib and E. Manes, "Machines in a Category: An Expository Introduction." *SIAM Rev.*, **16,** 1974, pp. 163–191.

30. ADJ = (J. A. Goguen, J. W. Thatcher, E. G. Wagner, J. B. Wright), "A Junction between Computer Science and Category Theory, I: Basic Concepts and Examples, Part 1." *IBM Research Report* RC 4526, T. J. Watson Research Center, September 1973.
31. ADJ, "A Junction between Computer Science and Category Theory, I: Basic Concepts and Examples, Part 2." *IBM Research Report* RC 5908, T. J. Watson Research Center, March 1976.
32. M. Arbib and E. Manes, *Arrows, Structures, and Functors: The Categorical Imperative.* Academic Press, New York, 1975.
33. S. Eilenberg and S. MacLane, "General Theory of Natural Equivalences." *Trans. Am. Math. Soc.* **58,** 1945, pp. 231–294.
34. D. Kan, "Adjoint Functors." *Trans. Amer. Math. Soc.*, **87,** 1958, pp. 294–329.
35. J. A. Goguen, "Minimal Realization of Machines in Closed Categories." *Bulletin of the American Mathematical Society*, **78,** No. 5, 1972, pp. 777–783.
36. F. W. Lawvere, "Adjointness in Foundations." *Dialectica*, **23,** 1969.
37. F. W. Lawvere, "Functorial Semantics of Algebraic Theories." *Proc., Nat'l Academy of Sciences*, **50,** 1963, 869 872.
38. F. Varela, "Not One, Not Two." *CoEvolution Quarterly*, Summer, 1976, pp. 62–67.
39. These functors correspond, in Figure 8, to the inclusion of subsystems within the boundary line of a system.
40. J. C. Smuts, *Holism and Evolution.* McMillan, New York, 1925.
41. E. Jantsch, "Evolution: Self-Realization Through Transcendence;" *Evolution and Human Consciousness*, edited by E. Jantsch and C. H. Waddington, Addison-Wesley, New York, 1976, pp. 37–63.
42. E. Laszlo, *An Introduction to Systems Philosophy.* Harper Colophon, New York, 1972.
43. K. Kosik, *La Dialectique du Concrete.* F. Maspero, Paris, 1969.
44. G. Radnitzky, *Contemporary Schools of Metascience.* Gateway Editions, Chicago, 1973.
45. J. A. Goguen, "Objects." *International Journal of General Systems*, **1,** No. 4, 1975, pp. 237–243.
46. C. Linde, "The Organization of Discourse." *The English Language: English in its Social and Historical Context*, edited by T. Shopen, A. Zwicks, and P. Griffin, 1977.
47. J. A. Goguen, "Complexity of Hierarchically Organized Systems and the Structure of Musical Experiences." *UCLA Computer Science Dept., Quarterly*, Oct. 1975, pp. 51–88; *International Journal of General Systems*, **3,** No. 4, 1977, pp. 237–251.
48. H. Maturana and F. Varela, *De Maquinas y Seres Vivos.* Editorial Universitaria: Santiago de Chile (English version: Biological Computer Lab. Report 9.4, U. of Illinois, Urbana, Ill. 1975), 1973.
49. F. Varela, H. Maturana and R. Uribe, "Autopoiesis: The Organization of Living Systems, Its Characterization and a Model." *Biosystems*, **5,** 1974, pp. 187–196.
50. H. Maturana, "Biology of Language: the Epistemology of Reality." *Psychology and Biology of Language*, a symposium in the memory of E. Lenneberg, Ithaca, Cornell U., May 1977.
51. A. N. Kolmogorov, "Logical Bases for Information and Probability Theory." *IEEE Trans. Inform. Th.*, **14,** No. 5, 1968, pp. 662 664.
52. A. N. Kolmogorov, *Mathematical Foundations of Probability Theory.* Chelsea 2nd English Edition, New York, 1956.

For biography and photograph of Professor Goguen and Dr. Varela, please see Vol. 1, *No.* 4, 1974, *and Vol.* 2, *No.* 1, 1975 *of this journal, respectively.*

nt. J. General Systems
979, Vol. 5, pp. 45–55

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

GROWING CELLULAR SYSTEMS: GENERATION OF GRAPHS BY PARALLEL REWRITING

ARISTID LINDENMAYER
Theoretical Biology Group, University of Utrecht, The Netherlands,

and

KAREL CULIK II

Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada.

(*Received July 6, 1977*)

Parallel graph generating systems have been constructed for the purpose of multidimensional simulation of developing multicellular organisms. They are extensions of the previously studied parallel string generating systems (L systems). They are discussed from the mathematical and the biological point of view.

INDEX TERMS Graph generating systems, graph grammars, parallel rewriting systems, cellular automata, polyautomata, L systems, developmental systems, multidimensional development.

NTRODUCTION

Cellular automata arrays[1] (polyautomata[2]) have een extensively studied in the past 30 years, artly with reference to computational problems switching circuits), and partly with biological notivation (nerve nets and self-reproduction). In ll of these cases, an interconnected collection of imilar subunits was assumed, each of the subnits usually not to be more complex than a finite utomaton. The behavior of the whole collection s then studied as a function of the behavior of he individual subunits.

We might say that problems concerning cellular automata arrays are representative *par excellence* of these of complex systems, while in ontrast to large dynamical systems, they have a ractable mathematical framework. In spite of the nany interesting results which have been obained, the above mentioned structures have certain shortcomings which makes them unsuitable or further biological applications, particularly in he area of development. In the case of the McCulloch–Pitts nerve net models, the problem that the size of the array remains constant; urthermore, that the entire array behaves as a ngle finite automaton. In the case of von Neumann's self-reproducing automata, the array laid out on a square grid and is allowed to row only at the margins (i.e. the active part of he array). While these structures are as powerful as Turing machines,[3] they are not suited to simulate in a realistic manner growing cells or multicellular organisms. The mathematical constructs presented in this paper overcome these shortcomings. They can represent interconnected systems of cells which are everywhere growing and are of any desired internal structure. Furthermore, they include as special cases the previously mentioned constructs, provided that they are extended to allow interactions.

PARALLEL REWRITING OF GRAPHS

All of the different kinds of cellular automata constructs share the property that the changes of state of the constituent subunits all take place simultaneously in discrete steps. This feature is also present in our systems, which is indicated by the term "parallel rewriting." We represent an interconnected system of cells by a directed graph with labelled nodes and edges. Each node represents a cell at a given time and its label corresponds to the state of the cell at that time. The whole life span of a cell (between the time it arises by division and the time it divides to give rise to new cells) corresponds thus to a sequence of labelled nodes. The labelled edges represent various kinds of connections between the cells. They can indicate either spatial directions, such as apical-basal or proximal-distal ones, or the

physiological nature of the cell boundaries, such as semipermeable membranes or cell walls with polarity. We allow only one of each type of edges to be present between a pair of nodes, and no edge from a node to itself. This rules out certain cellular configurations, but they are so rare that these restrictions appear to be desirable.

We plan to describe multicellular organisms throughout an extended period of time and not only at one stage. We approach this goal by parallel rewriting of our graphs. This means that we have to introduce rules by which a new graph can be produced from a previous one. Using such a system of rules, one obtains either a sequence of graphs (in their order of development), or a set of graphs (without an ordering) which is usually called a "graph language."

Graphs may also be rewritten sequentially (for a recent review of graph grammars, also called "web grammars," see reference[4]). In these sequential graph rewriting systems, a particular subgraph is replaced by another subgraph at each time step. These are useful for such applications as modifications of flow charts, but not for our purposes, because we want to allow simultaneous changes throughout the graph as it happens in real organisms. It should be obvious that it is not possible to convert a sequential graph rewriting system into a parallel one without completely redefining the whole procedure.

The advantage of using labelled graphs rather than strings of symbols in a generating system is that one is not restricted to one-dimensional simulations of growing organisms. However, graphs are by themselves neither two- nor three-dimensional structures, they are simply sets of ordered pairs of nodes. Planar graphs can be easily and uniquely interpreted as two-dimensional maps, which in turn can be directly applied to figures obtained from developing organisms. Parallel map generating systems have been proposed.[5,6,7] Because they are applicable only to two-dimensional development and because of easy convertability of maps and planar graphs, they do not offer any advantage over the parallel graph generating systems.

Parallel graph rewriting was introduced by us[8] following certain suggestions of Mayoh.[5] Other studies of these and related systems have since been published,[9-16] extending and modifying some of our assumptions. Although the step from graphs to three-dimensional cellular packed structures is not as obvious as is the case for two-dimensional ones (tesselations), the graph approach remains at the moment the most promising one for three-dimensional simulation of development.

DEFINITION OF PGOL SYSTEMS

A parallel graph rewriting system in our sense consists of a set of node labels, a set of edge labels, a set of production rules, a set of connection rules, and a starting graph. A production rule specifies a graph which can be substituted for a certain node in the previous graph. In a graph rewriting system with interactions, the substitutions might depend on the surroundings of the given node (context-sensitive rewriting), while in a system without interactions, the substitutions are determined only by the label of the node being substituted (context-free rewriting). We shall only consider here the latter kinds of systems.

After the substitutions for the nodes of the previous graph have been completed, we apply the connection rules in order to interconnect the individual subgraphs. One obviously has many choices as to the kinds of connection one wishes to allow. Since we wish to preserve the neighborhoods of cells and their descendants in the course of development (cells are not supposed to move past each other), we restrict our connection rules in such a way that they prescribe connections only between those nodes which are produced from previously connected (mother) nodes. Furthermore, in our systems, the label of the directed edge between the (oriented) pair of mother nodes in the previous graph is the determining factor for the connections. This label determines how a given (oriented) pair of substituted graphs (daughter graphs) should be interconnected. These interconnections are therefore performed independently on each pair of daughter graphs on the basis of the label of each edge between their mother nodes. This is a simple solution[5] to the problem of covering, by finitely many connection rules, all possible situations, even that of nodes having unboundedly many edges. While weaker than the assumption that all edges on a mother node should affect the interconnections between its daughter graph and the surrounding graphs, our assumption is sufficient to our purpose and is defensible on cytological grounds.

Using the production and connection rules, we write step-by-step a particular starting graph. ur systems can be deterministic or non-terministic. In the deterministic case, each next ep in the rewriting is determined uniquely, so a ngle sequence of graphs is obtained from the arting graph. In the non-deterministic case, any different derivation sequences can occur. e can then consider the collection of all graphs at is produced (the graph language).

Productions may be formulated in such a way at a node is replaced by the empty graph, i.e., is erased. Systems in which no erasing is lowed are called "propagating." We consider in is paper only propagating systems.

Parallel rewriting systems on one-dimensional llular arrays (represented by strings of symbols) ve been called "OL systems" if no interaction kes place among the cells, and "IL systems" if ere is interaction.[17-20]. Our systems are an tension of these string rewriting systems, and e therefore called "propagating graph OL sys-ms," abbreviated as "PGOL systems."

In the following, we give an informal definition PGOL systems; for the detailed definition see ur previous paper.[8]

As stated before, we are dealing with directed aphs which have labelled nodes and labelled ges. Each graph is to have a special node, the vironmental node, labelled by e. Edges going or from the environmental node shall be called utside" edges, and the others "inside" edges. nce we are not interested in naming individual des, we shall only be concerned with isomor-ic classes of such graphs, in the following lled "abstract e-graphs." Whenever we speak of "node" or an "edge" of an abstract e-graph, we tually mean a node or an edge of one of the ncrete graphs in this class. This is done only in tuations where the choice of the concrete graph irrelevant.

An abstract e-graph over (A,E) is an isomor-ic class of graphs, each with an environmental de labelled $e(e\notin A)$, all the other nodes labelled elements of A, and edges labelled by elements E. The family of all abstract e-graphs over $,E)$ is denoted by $(A,E)_*$. The empty abstract graph (the abstract e-graph with a single node belled e and no edges) is denoted by λ, and the $(A,E)_*-\{\lambda\}$ is denoted by $(A,E)_+$. In our agrams, the outside edges will be shown as free rows and the environmental node will not be own. We call the outside edges "hands."

We want to define now the coalescence of two abstract e-graphs into one by adding new edges. This will be done with the help of "stencils." A stencil is an abstract e-graph with a bipartition of its nodes. In consequence, each stencil has two sets of nodes, and we call these the "source" and the "target" nodes.

An ordered pair of abstract e-graphs (from now on we shall simply call them "graphs") α,β can be joined (coalesced) according to the stencil γ into a new graph δ, denoted as $\alpha\overset{\gamma}{\rightarrow}\beta$, in the following way. A pair of nodes n_1,n_2 from α and β may be joined by a directed edge labelled G only if n_1 has outgoing hand labelled G and n_2 and incoming hand labelled G. Which of such pairs of nodes are connected is determined by the stencil γ in such a way that after all new connections are made γ is a subgraph of δ so that:

i) the source nodes of γ are all in α,

ii) the target nodes of γ are all in β,

iii) all the new edges between nodes of α and β are in γ.

For a stencil γ let γ_S, γ_T denote the subgraphs induced by its source and target nodes respectively. Let C be a set of stencils, and γ be in C. We say that γ is C-maximal for α,β if γ is applicable to α,β and if there is no ξ in C such that ξ is applicable to α,β; γ_S is a subgraph of ξ_S and γ_T is a subgraph of ξ_T.

We can now define a PGOL system as follows. A PGOL system G is a quintuple $\langle A, E, P, C, S\rangle$, where

A is a finite set of node labels (node alphabet),
E is a finite set of edge labels (edge alphabet),
P is a finite subset of $A\times(A,E)_+$ (set of production rules),
C is a finite subset of $E\times(A,E)^S_+$, where $(A,E)^S_+$ is the set of stencils over $(A,E)_+$ (C is the set of connection rules), and
S is an element of $(A,E)_+$ (the starting graph).

Production and connection rules are written in the form $a\mapsto\alpha$. The set P must be complete, in the sense that for each a in A there must be an α so that $a\mapsto\alpha$ is in P.

For abstract e-graphs γ, δ in $(A,E)_+$ we write $\gamma\Rightarrow\delta$ to mean that γ is rewritten into δ, if δ can be obtained from γ in the following way:

i) Each node of γ (mother node) is replaced by a graph (daughter graph) so that if node x

which is labelled a is replaced by graph σ_x then $a \mapsto \sigma_x$ must be a production in P.

ii) The graphs obtained in (i) are interconnected so that for each edge, say q, going from node x to node y of γ, certain interconnections (possibly none) are added between the graphs σ_x and σ_y of δ which are obtained from mother nodes x and y. We stress that these interconnections are made for each edge between the mother nodes independently of their other edges. If edge q has label H then we choose any connection rule $H \mapsto \beta$ such that β is a C-maximal stencil applicable to σ_x and σ_y, and we join σ_x and σ_y according to β so that graph $\sigma_x \overset{\beta}{\rightarrow} \sigma_y$ is obtained (disregarding the new edges possibly contributed by other edges between mother nodes x and y). If there is no rule $H \mapsto \beta$ applicable to σ_x and σ_y then no edges are created between them.

iii) The hands (both incoming and outgoing) of δ are obtained as follows. If a mother node x of γ has an incoming (outgoing) hand labelled H, then all incoming (outgoing) hands labelled H of graph σ_x of δ are retained; the others disappear.

For a more rigorous, but also more complex definition of the relation $\Rightarrow$ see our previous paper.[8]

The reflexive and transitive closure of the relation $\Rightarrow$ is written as $\overset{*}{\Rightarrow}$, denoting the rewriting of one graph into another in some finite number of steps.

The graph language generated by the PGOL system G, $L(G)$, is defined as the set $\{\alpha : S \overset{*}{\underset{G}{\Rightarrow}} \alpha\}$, where S is the starting graph of G.

Previously we have also shown[8] that using stencils, we can define a mapping, representing coalescence of graphs, from sets of graphs into graphs. Repetitive applications of such mappings can be expressed by recurrent formulas, like those defined for strings.[21] Sets of such formulas can be used as alternative generating systems for graph languages. These have been shown by us to be stronger than PGOL systems, i.e., they can generate graph languages which cannot be generated by PGOL systems, but not *vice versa*. Special cases of the recurrence systems are the locally catenative systems[22] which have been particularly useful in biological applications.[23]

EXAMPLES OF PGOL SYSTEMS

1) A developmental pattern which is very often encountered in algae, mosses and ferns can be expressed by a simple PGOL system. We show the pattern in Figure 1, in a number of developmental stages. Periodically the apical cell (labelled a) divides into a new apical cell and a basal cell (labelled b). The basal cells remain undivided; they are shown here to undergo changes of state from b to c, and then to d, which are necessary to obtain simple connection rules. Thus the production rules are:

$$a \mapsto \quad a^{\nearrow} \rightarrow b \rightarrow$$
$$b \mapsto \rightarrow c \rightarrow$$
$$c \mapsto \rightarrow d \rightarrow$$
$$d \mapsto \rightarrow d \rightarrow$$

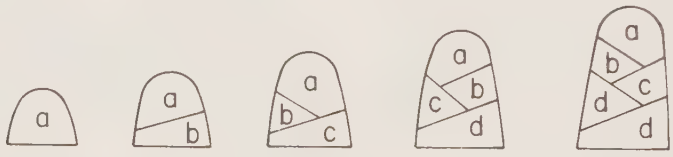


FIGURE 1

Note that we use $\mapsto$ type arrow between the left and right side of a production (or connection) rule, while a $\rightarrow$ type arrow is used to indicate directed edges of a graph. The node alphabet in this case is $\{a, b, c, d\}$, and the edge alphabet consists of a single symbol, #, which in such a case is not shown on the arrows.

The connection rules are as follows:

$$\# \mapsto \begin{matrix} & & a_S & & \\ & \searrow & & & \\ & & & \searrow & c_T \\ & \nearrow & & & \\ & & b_S & & \end{matrix}$$

$$\# \mapsto x_S \rightarrow d_T \quad \text{for } x \in \{b, c, d\}$$

Beginning with a starting graph consisting of a symbol a, the first four generation steps yield the graphs shown in Figure 2. We see that a ladder-like chain of cells is produced, corresponding to the map representation of these structures in Figure 1. The substitution of daughter graphs and their interconnections are extremely simple in this case.

In order to illustrate that the same ladder-like structures, but with different node labels and different edge orientations, can be produced by another system, we show the following rules. In this case we need only two node labels, a and b,

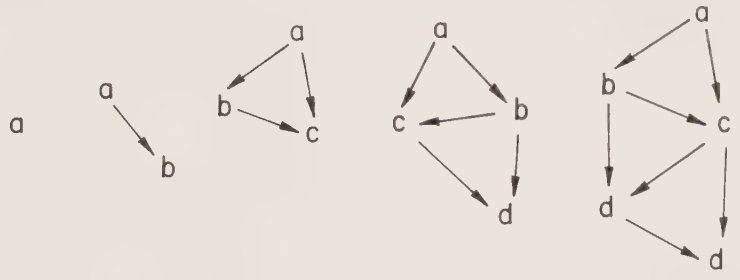


FIGURE 2

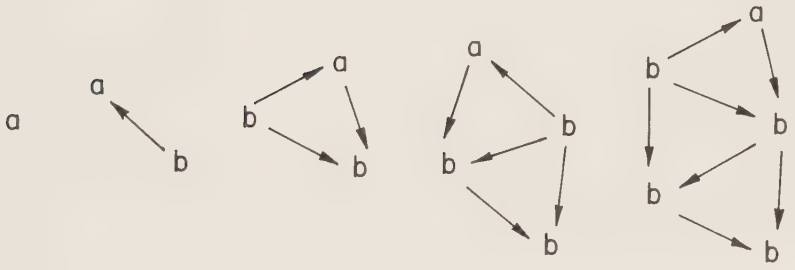


FIGURE 3

and a single edge label, #. The production rules are:

$$a \mapsto \overset{\nearrow}{a} \leftarrow \overset{\nearrow}{b}$$

$$b \mapsto \rightarrow b \rightarrow$$

and the connection rules:

$$\# \mapsto b_S \begin{matrix} \swarrow a_T \\ \nwarrow b_T \end{matrix}$$

$$\# \mapsto b_S \rightarrow b_T$$

Note that the arrow in the first production points from b to a in this case, and that the arrows in the first connection rule point from target to source nodes. This effectively means a turning around of the polarity of the cellular connections in the second step after a basal cell is generated. At the price of this assumption, the number of node labels was reduced from four to two as compared to the previous model. Figure 3 shows the sequence of graphs generated by the second model.

A more elaborate example of this kind of development, that of the leaf of the moss *Phascum cuspidatum*, is shown by Nagl,[9,10] based on a map system.[24]

2) Another commonly encountered developmental pattern in plants is the case of the propagation of polygonal patterns of cells on a growing cylindrical of conical apex. We show here how a particular polygonal pattern, in this case pentagonal, can be propagated by simple production and connection rules. We show these conical arrays from above, in the form of concentric pentagonal rings.

Two node labels are needed: a and b, and a single edge label: #. The production rules are:

$$a \mapsto \rightarrow b \rightarrow a \rightarrow$$

$$b \mapsto \rightarrow b \rightarrow$$

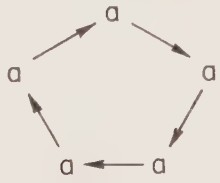
and the connection rules are:

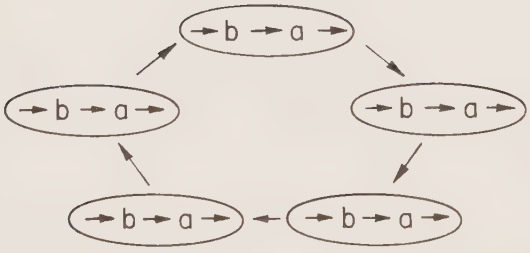
$$\begin{array}{ccc} & b_S & b_T \\ \# \mapsto & \downarrow & \nearrow \\ & a_S & \rightarrow a_T \end{array}$$

$$\# \mapsto b_S \rightarrow b_T$$

We assume a pentagonal array of nodes labelled a as the starting graph. For the sake of exposition we show in Figure 4 the first generation step in two stages, first the substitutions of the daughter graphs, and then their proper interconnections forming the new graph.

We see that by our maximality condition, the second stencil cannot be applied in this step.





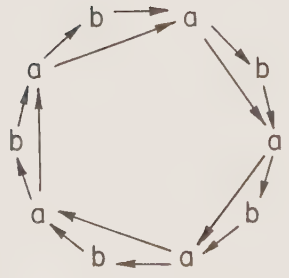
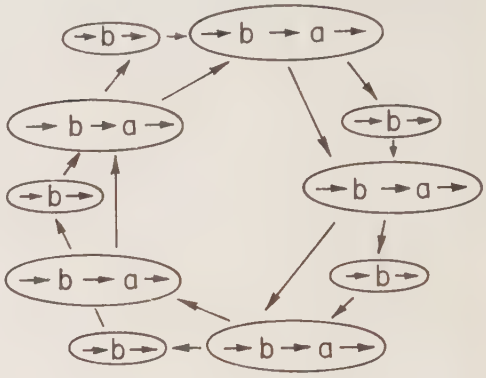


FIGURE 4

The next step is shown in Figure 5. Here the second stencil can be applied when one of the daughter graphs are of the kind

$$\rightarrow b \rightarrow$$

We see that a regularly connected series of five-membered arrays of b's are produced in each generation step. These are described in the botanical literature as five-membered whorls of



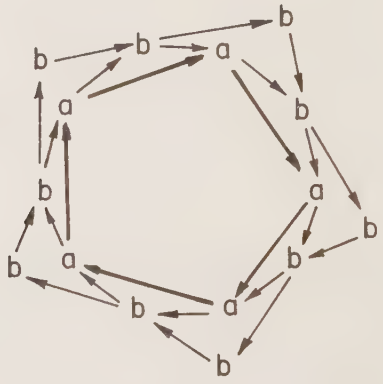


FIGURE 5

cells or organs (like leaves) on a growing axis. They may be considered as simple phyllotactic patterns. With modifications of the production and connection rules, one can easily get a large variety of these patterns. It should also be noted that the number of whorl members in the above system depends only on the starting configuration.

In these examples, no hands are present on the starting graphs, thus none of the graphs generated from them have hands. Hands are primarily necessary to govern the interconnections between the daughter graphs, the mothers of which were connected. They can be used effectively to reduce the number of edge labels needed. If it is not thought desirable to use hands in this way, then one can simply assume that hands (incoming and outgoing) of all possible labels are present on all nodes.

3) A class of developmental patterns very important for description of plants is that of branching structures. These are different from the structures called "trees" in graph theory in that there are two fundamentally different kinds of branches that can arise from a node: branches which continue straight from the node and branches which are inserted at an angle to the direction of the previous segment. In order to express this distinction, we use two edge labels: D for straight and B for branching edges. A variety of branching patterns have been described previously,[18,34] and we shall only show here very briefly how two well-known patterns, the monopodial and the sympodial branching patterns can be generated by our graph systems.

For the monopodial pattern we have the following production rules:

$$a \mapsto \begin{array}{l} \overset{B}{\nearrow} b \\ \overset{D}{\rightarrow}\ b \\ \underset{D}{\searrow} a \end{array}$$

$$b \mapsto \begin{array}{ccc} & \uparrow_B & \\ \overset{D}{\rightarrow} & b & \overset{D}{\rightarrow} \\ & \uparrow_B & \end{array}$$

and connection rules:

$$D \mapsto b_S \overset{D}{\rightarrow} b_T$$

$$B \mapsto b_S \overset{B}{\rightarrow} b_T$$

A derivation sequence is shown in Figure 6.

For the sympodial pattern we need to change only the first production rule in the following way:

$$a \mapsto \begin{array}{l} \overset{B}{\nearrow} a \\ \overset{D}{\rightarrow}\ b \\ \underset{D}{\searrow} b \end{array}$$

The derivation sequence is then the one shown in Figure 7.

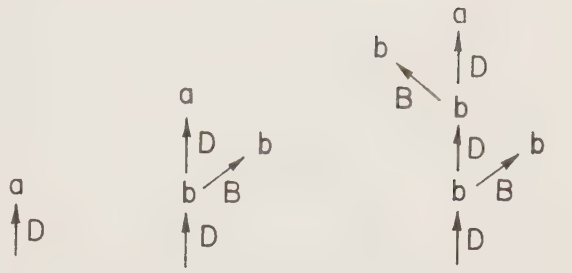


FIGURE 6

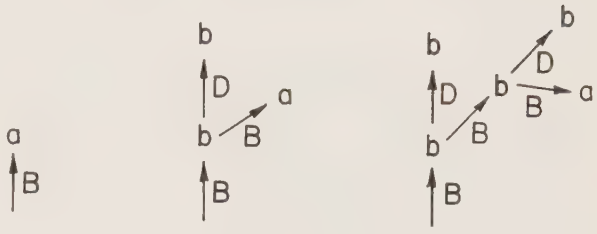


FIGURE 7

These branching patterns can also be generated by OL systems with a coding that transforms branching structures into strings of symbols with parentheses. The advantages of graph generating systems in this case is that one can dispense with the additional coding step which imposes a certain restriction on the types of structures that can be generated.

Since the starting graphs of these systems have incoming hands labelled D and B respectively, the derived graphs also have these hands.

BIOLOGICAL PERSPECTIVES

The biological basis and motivation of developmental models of this type have been discussed before.[19,24] The choice of discrete models, in which we are dealing with distinct spatial units, the cells, with discrete time steps, and with finitely-many cellular states, was defended there in some detail. But previously only one-dimensional cellular arrays were considered, where the manner of connections presented no particular problems, not even in the case of branching structures. Now that we have started to consider multidimensional arrays, we should examine the foundations and the biological justification of the various additional control mechanisms which we were forced to introduce in the definition of

PGOL systems. A detailed examination cannot be presented here; it will be done in more biologically oriented articles. It is possible that some of the features introduced by us are not the best suited for biological interpretation and will eventually be replaced by others. It is, however, our present opinion that all aspects of these systems are compatible with the facts of developmental biology, cell physiology and morphology. These systems were, after all, constructed in close cooperation between a biologist and a mathematician, explicitly as developmental models.

The primary control factor in our systems, as far as the connections between pairs of daughter graphs are concerned, are the directed, labelled edges between pairs of mother nodes. We have already alluded to the question of whether this control should be by pairs of mother nodes, or by the whole neighborhood of a mother node. As long as it is assumed that the direction in which cells divide in tissues is mainly controlled by simple gradients, our type of control would be sufficient. The fact that we are using directed edges is also geared to the gradient-type mechanisms. Of course, if there is no need for these kinds of oriented mechanisms, one could use double edges in the models.

The edges between the mother nodes determine the stencils which can be used to produce the new connections. The stencils must be applicable to the daughter graphs, i.e., the right kinds of nodes and the right kinds of edges must be present in a proper arrangement in both the source and target graphs. This means that all of these elements can have an influence on the connections to be made. These are precisely the influences that one needs if one wishes to simulate the effect of a neighbor cell (e.g. by producing an active substance) on the direction of a particular cell division. If no such control is needed, then one can adopt uniform connection rules, i.e., stencils which are of the same form for all edge labels.

Finally, hands also have an effect on the making of new connections. Their biological interpretation is straightforward: they indicate the cells which are on the outside of an array. If the array is a daughter graph, then the cells with hands are the ones which are available for connections. But since hands are directed half-edges, they have a function in also determining the direction in which the daughter graphs are to be connected and nodes which are to be employed. Thus they can be very useful in simulating polar development, or development in which certain directed cell files have to be formed. In addition, when graphs are generated by recurrence systems, hands are also useful to indicate the outside nodes on the larger and larger graphs which are joined together to form the whole structure. On the other hand, if, in a particular application, no hands are required, it is easy to exclude them from exerting control by assuming that they are present on every node.

The use of these models for developmental biology is of two sorts. On the one hand they can be used as simulation tools, frameworks within which detailed simulations of particular developing organisms can be carried out. While OL and IL systems have been employed in this way,[25-28] no such work has yet been done with the help of PGOL systems. There are, however, models of the development of snail embryos, chicken limbs and insect eyes[29-33] which represent discrete simulations in two- or three-dimensional arrays, and are related to PGOL systems or to similar graph generating systems with interactions.

On the other hand, these models can also serve to guide our intuition concerning the underlying mechanisms of development such as gradients, polarity, organizers, morphogenetic fields, and their roles in controlling directed cell divisions and directed growth. The major insight which resulted from the study of OL and IL systems has been the clear cut distinction found between the developmental patterns that can be obtained without cell interactions as against those patterns for which cell interaction is necessary. Now that the theory has yielded this distinction, it is easy to recognize it in a wide variety of organisms. Similarly, it is hoped that the study of PGOL systems, and their counterparts with cell interactions (PGIL systems), will also yield insights into the mechanisms which are necessary and/or sufficient for the production of certain developmental patterns. That the basic distinction between graph systems with and without interaction is just as valid as it is in the string systems, has already been confirmed.[8] Further important distinctions may be found by studying the patterns which can be generated with a single node label or with a single edge label, or with undirected edges, or with uniform stencils, or with uniform distribution of hands.

OMMENTS ON THE CHOICE OF PGOL YSTEMS AS TOOLS FOR THE PARALLEL ENERATION OF GRAPHS

n the previous sections, we have discussed the uestion of what kind of graph rewriting is eeded for representing the essential aspects of eveloping organisms, in particular plants. In the ollowing, we summarize the main assumptions hich should be satisfied by generating systems uitable to the above purpose.

i) At every step of rewriting each (mother) ode of the "old" graph is to be replaced by a on-empty subgraph of the "new" graph. The eplacement should depend only on the label of he mother node. In this sense, we are dealing ith an interactionless or context-free rewriting.

ii) Nodes of the "new" graph may be con-ected only if they have the same mother node, r if their mother nodes were connected in the old" graph.

iii) The connection of two nodes with different nother nodes depends only on the label and rientation of the edges between their mother odes and on the subgraphs by which the mother odes were replaced.

Our intention in introducing PGOL systems as been not only to provide a framework within hich every rewriting system which satisfies con-itions (i)–(iii) can be presented, but more impor-antly, to provide a tool which would be re-atively easy to use when describing real biologi-al or other objects. The price to pay for this atter requirement is that the precise mathemati-al definition of PGOL systems[8] is rather com-plicated. This is analogous to the situation, say, n programming languages: clearly a high-level rogramming language like ALGOL is easy to se but its formal definition (syntax and seman-ics) is rather complex when compared to, say, ssembly code. We have a less drastic but anal-gous situation in our case. We will show that here is a special subclass of "complete" PGOL ystems which retains all of the generation power f PGOL systems but has a much simpler formal efinition. However, to use only complete PGOL ystems in practice would be more cumbersome.

A PGOL system $G=\langle A,E,P,C,S\rangle$ is called *omplete* if the source (target) part of each stencil n C is the right-hand side of a production in P, nd if for each oriented pair of right-hand sides f productions there is at least one stencil in C.

We will not repeat the formal definition of the relation $\Rightarrow$ for this special case, but it should be clear that it would be much simpler to define it rigorously than in the general case. Here, for every combination of "right-hand sides" we can find in a "list" the instructions for new connections. There is no problem which stencil fits a particular graph, and there is no need for the maximality condition. We simply use the list of one or more connection rules (stencils) for each edge label and pair of "right-hand sides." We will show that complete PGOL systems retain all generative power of PGOL systems, i.e., we prove that they are a normal form for PGOL systems (a similar theorem was stated previously in terms of Nagl's parallel graph systems[16]).

Normal form theorem. For each PGOL system $G=\langle A,E,P,C,S\rangle$ there exists a complete PGOL system $\bar{G}=\langle A,E,P,\bar{C},S\rangle$ such that $\alpha\underset{\bar{G}}{\Rightarrow}\beta$ if and only if $\alpha\underset{G}{\Rightarrow}\beta$ for all α, β in $(A,E)_+$.

Proof We construct $\bar{C}$ (the completion of C) as follows. For each B in E and each pair x, y of right-hand sides of productions in P we include in $\bar{C}$ all (finitely many) connection possibilities allowed by stencils in C, each of them described by a stencil with source part x and target part y. Clearly, if $\alpha\underset{\bar{G}}{\Rightarrow}\beta$ then $\alpha\underset{G}{\Rightarrow}\beta$. If no right-hand side of a production in P is a subgraph of another right-hand side, then the reverse implication is also obvious. In the case that one right-hand side is a subgraph of another one, then the reverse implication follows from the maximality condition, i.e., from the condition that a stencil may be applied only if it is a $\bar{C}$-maximal stencil.

We may add the following comments. The use of complete PGOL systems might require a much larger number of stencils than the equivalent non-complete systems. For practical applications, when we are composing rules for a real developmental process, it is more convenient to allow the use of arbitrary (possibly non-complete) PGOL systems. However, in the case of a computer implementation of PGOL systems, it might be useful to let the computer first perform the construction indicated in our normal form theorem and convert the arbitrary PGOL system into a complete one, before it begins with carrying out computations involved in the generating steps. It is easy to see that this would lead to a faster program, but at the price of a higher memory requirement.

Even faster programs, but with still higher memory requirements, could be obtained if we let

the graph be pre-computed, which is to be substituted for each triple (a, H, b), where a, b are node labels and H is an edge label. In this way we can merge the production and connection rules into rules of the form $(a, H, b) \mapsto \alpha$. These kinds of rules represent, of course, a very redundant way of writing instructions because to the same mother node there might be many such triples applicable. In that case, the right-hand sides of the original production for a would be repeated many times.

In the case of non-deterministic PGOL systems, this would probably not be a very efficient way to program graph generation because the overlapping parts of the right-hand sides of instructions for triples would have to be matched. Parallel graph generating systems where the definition is based on such instructions for triples (called "handles") have been considered by Ehrig and Rozenberg.[13] We do not feel that the approach based on handles gives a convenient tool for describing parallel graph generation, since by the above discussion it can be seen that in most cases it would yield redundant and much larger descriptions than the equivalent PGOL systems.

Parallel graph generating systems which allow more general connection rules than specified in our conditions (ii) and (iii) have been studied by Nagl[9-11] and Grötsch and Nagl.[16] For certain biological applications, these extensions are necessary.

We have mentioned before that our production rules are of a context-free nature. This statement is unquestionably true, but it is less clear whether an entire PGOL system, including connection rules, can also be considered context-free. There is no generally accepted definition for context-freeness for graph languages, so we can only approach this subject intuitively.

Context-sensitiveness should intuitively imply communication among the parts of the structures being generated. On the other hand, it should be typical for context-free generation that it would have a "local" property, an extension of the following property of context-free generation of strings. Namely, what can be generated from the string uv is the concatenation of what can be generated separately from strings u and v. If, instead of concatenation we use the more complicated concept of graph coalescence, this local property holds for PGOL generation.

On this property is based the proof of an important result proven in our previous paper,[8] namely, that the subgraph problem is decidab for PGOL languages. This means that given PGOL system G and a graph α, it is recursive decidable whether α is a subgraph of a memb of the language of G.

While these considerations argue for contex freeness of PGOL systems, others militate again this conclusion. For further discussion of th point see ref. 36.

REFERENCES

1. A. W. Burks, editor, *Essays on Cellular Automa* University of Illinois Press, Urbana, 1968.
2. A. R. Smith III, "Introduction to and survey of p lyautomata theory." In: Ref. 35, 1976, pp. 405–422.
3. M. A. Arbib, "Automata theory in the context of theore cal embryology." In: *Foundations of Mathemati Biology*, edited by R. Rosen, Academic Press, New Yo 1972, Vol. 2, pp. 141–215.
4. A. Rosenfeld, "Array and web grammars: An overvie In: Ref. 35, 1976, pp. 517–529.
5. B. H. Mayoh, "Multidimensional Lindenmayer system In: *L Systems*, edited by G. Rozenberg and A. Saloma Lecture Notes in Computer Science, Vol. 15, Sprin Verlag, Heidelberg, 1974, pp. 302–326.
6. B. H. Mayoh, "Another model for the development multidimensional organisms." In: Ref. 35, 1976, pp. 46 485.
7. J. W. Carlyle, S. A. Greibach and A. Paz, "A tw dimensional generating system modelling growth by nary cell division." In: *15th Annual Symp. on Switch and Automata Theory, Proceedings*, IEEE, New Yo 1974, pp. 1–12.
8. K. Culik II and A. Lindenmayer, "Parallel graph gener ing and graph recurrence systems for multicellular velopment." *Int. J. General Systems*, **3,** No. 1, 1976, p 53–66. (A preliminary version has appeared as Techni Report CS-74-22, Faculty of Mathematics, University Waterloo, Nov. 1974).
9. M. Nagl, "Graph Lindenmayer-systems and language In: *Arbeitsberichte Inst. für math. Maschinen u Datenverarbeitung*, Univ. Erlangen-Nürnberg, **8,** No. Jan. 1975, pp. 16–63.
10. M. Nagl, "On a generalization of Lindenmayer-systems labelled graphs." In: Ref. 35, 1976, pp. 487–508.
11. M. Nagl, "Graph rewriting systems and their applicati in biology." In: *Mathematical Models in Medicine*, edit by J. Berger, W. Bühler, R. Repges and P. Tautu, Lectu Notes in Biomathematics, Vol. 11, Springer Verl Heidelberg, 1976, pp 135–156.
12. H. Ehrig and H. J. Kreowski, "Parallel graph grammar In: Ref. 35, 1976, pp. 425–442.
13. H. Ehrig and G. Rozenberg, "Some definitional sugg tions for parallel graph grammars." In: Ref. 35, 19 pp. 443–468.
14. P. J. A. Reusch, "Generalized lattices as fundamentals retrieval models, multidimensional developmental syste and the evaluation of fuzziness." In: *Informatik Berich* Univ. Bonn, No. 10, July 1976, 205 pp.

5. H. D. Weckmann, "Mehrdimensionale Parallele Ableitungssysteme—basierend auf markierten Hypergraphen—und ihre Anwendungsmöglichkeiten." In: *Informatik Berichte*, Univ. Bonn, No. 11, Sep. 1976, 66 pp.
6. M. Nagl, "On the relation between graph grammars and graph Lindenmayer-systems." In: *Arbeitsberichte Inst. für math. Maschinen und Datenverarbeitung*, Univ. Erlangen-Nürnberg, Vol. 9, No. 1, 1976, pp. 3 32; see also technical report by E. Grötsch and M. Nagl, "Comparison between explicit and implicit graph L-systems," from the same institute, 1977.
7. A. Lindenmayer, "Mathematical models for cellular interactions in development, Parts I and II," *J. Theor. Biology*, **18,** 1968, pp. 280–315.
8. A. Lindenmayer, "Developmental systems without cellular interactions, their languages and grammars," *J. Theor. Biology*, **30,** 1971, pp. 455–484.
9. A. Lindenmayer, "Developmental algorithms for multicellular organisms: A survey of L systems." *J. Theor. Biology*, **54,** 1975, pp. 3–22.
10. G. T. Herman and G. Rozenberg, *Developmental Systems and Languages.* North-Holland/American Elsevier, Amsterdam/New York, 1975.
11. G. T. Herman, A. Lindenmayer and G. Rozenberg, "Description of developmental languages using recurrence systems," *Math. Systems Theory*, **8,** 1975, pp. 316–341.
12. G. Rozenberg and A. Lindenmayer, "Developmental systems with locally catenative formulas," *Acta Informatica*, **2,** 1973, pp. 214–248.
13. H. B. Lück, "Elementary behavioural rules as a foundation for morphogenesis," *J. Theor. Biology*, **54,** 1975, pp. 23–34; see also H. B. Lück and J. Lück, "Cell number and cell size in filamentous organisms in relation to ancestrally and positionally dependent generation times." In: Ref. 35 1976, pp. 109–124.
14. A. Lindenmayer, "Developmental systems and languages in their biological context." In: Ref. 20, 1975, pp. 1–40.
15. R. Baker and G. T. Herman, "Simulation of organisms using a developmental model, Parts I and II," *Int. J. Bio-Medical Computing*, **3,** 1972, pp. 201–215, 251–267.
16. J. Lück, "Complémentarité apico-basale dans l'organisation des tissus végétaux." Thèse docteur ès sciences, Univ. Aix-Marseille III, June 1977.
17. D. Frijters, "Control mechanisms and computer descriptions of inflorescence development." Proefschrift doctor wiskunde en nat. wet., Rijksuniv. Utrecht, May 1976.
18. D. Frijters, "An automata-theoretical model of the vegetative and flowering development of *Hieracium murorum* L." *Biological Cybernetics*, **24,** 1976, pp. 1 13.
19. C. P. Raven and J. J. Bezem, "Analysis of pattern formation in gastropods by means of computer simulation." In: Ref. 35, 1976, pp. 139–145.
20. J. J. Bezem and C. P. Raven, "Some algorithms used in the simulation of embryonic development." In: Ref. 35, 1976, pp. 147 158.
21. D. A. Ede and O. K. Wilby, "Analysis of cellular activities in the developing limb bud system." In: Ref. 35, 1976, pp. 15–24.
22. O. K. Wilby and D. A. Ede, "Computer simulation of vertebrate limb development the effect of cell division control patterns." In: Ref. 35, 1976, pp. 25 37.
23. R. Ransom, "Computer analysis of cell division in *Drosophila* imaginal discs: Model revision and extension to simulate leg disc growth." *J. Theor. Biology*, **66,** 1977, pp. 361–377.
24. D. Frijters and A. Lindenmayer, "Developmental descriptions of branching patterns with paracladial relationships." In: Ref. 35, 1976, pp. 57–73.
25. A. Lindenmayer and G. Rozenberg, editors, *Automata, Languages, Development.* North-Holland, Amsterdam, 1976.
26. K. Culik II and D. Wood, "A mathematical investigation of parallel graph OL systems." Research Report CS-78-36, Faculty of Mathematics, Univ. of Waterloo, Aug. 1978.

For Dr. Lindenmayer's biography and photograph, please see Volume 2, No. 4, p. 190 of this journal; for Dr. Culik's biography and photograph, please, see Volume 3, No. 1, p. 66 of this journal.

Int. J. General Systems
1979, Vol. 5, pp. 57–61

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

Letters to the Editor

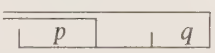
A COMMENT ON VARELA'S CALCULUS FOR SELF-REFERENCE

Sir—In Francisco Varela's "Calculus for Self-Reference"† there is an error in the proof of Theorem 10, the principal formal result of the paper. The error does not invalidate the final result.

By a self-referential or re-entrant expression is meant one defined by an equation of the form

$$f = h(f)$$

The formalism used by Brown and Varela to indicate re-entry requires that the expression $h(f)$ consists of a cross and a sub-expression pervading it. Only when $h(f)$ has this form can the points of re-entry be indicated by linking them to the lower tip of the cross, as when



is used to represent the solution of

$$f = \overline{pf \mid qf}$$

The restriction on $h(f)$ places no ultimate restriction on the expressions which can be represented in the calculus, since by Varela's Proposition 1,

$$h(f) = \overline{\overline{h(f)}}$$

If, therefore, $h(f)$ is not originally of the required form an equivalent expression meeting the requirement can be obtained by adding two crosses.

However, the restriction is important when it is necessary to consider the depth at which re-insertion is made in an expression. Varela appears to have overlooked this in deriving his Theorem 6, on page 14 of his paper. Theorem 6 states that, for any given expression, an equivalent expression not more than two crosses deep can be derived. The proof depends on application of his Proposition 4:

Unfortunately the r.h.s. of P4 is not roofed over by a single cross and is not admissible as a re-entrant expression. It is not permissible to change it to the required form by placing two crosses over it since the requirement of the Theorem is that the re-entry be not more than two crosses deep.

The only consequence of modifying Theorem 6 to allow re-entry at greater depth is that the number of forms of re-entry which have to be considered in the proof of Theorem 10 is rather more than the three considered by Varela. However, the extra forms can be treated in essentially the same way as these three are treated, and the validity of Theorem 10 is not affected.

Theorem 10 can, in fact, be proved in an alternative and rather simple-minded way. The alternative is no substitute for Varela's full treatment, but may provide additional insights.

Let the three basic values of the calculus, *null*, ☐ and ⌝, be represented by the integers 1, 2 and 3 respectively. It is then useful to consider the *value vectors*, or ordered sets of values which may be assumed by sub-expressions as the overall expression takes each of the values (1, 2, 3) in turn. Clearly the *value vector* for a point where re-entry is made is (1, 2, 3), while indications of *null*, ☐ and ⌝ have respective value vectors (1, 1, 1), (2, 2, 2) and (3, 3, 3). Thus there are four primary value vectors.

The calculus depends on two operations, namely the juxtaposition of two sub-expressions to produce a value determined by the rule of dominance, and the crossing of a sub-expression. The corresponding operations on value vectors are

†F. Varela, "A calculus for self-reference". *Int. J. General Systems*, **2**, pp. 5–24, 1975.

obvious. Juxtaposition produces a new vector whose entries are the suprema of the pairs of entries in corresponding positions, so that (1, 2, 3) and (2, 2, 2) would give rise to (2, 2, 3). Crossing corresponds to the substitution $1 \rightarrow 3$, $2 \rightarrow 2$, $3 \rightarrow 1$, so that (2, 2, 3) would give rise to (2, 2, 1).

It is clear that Theorem 10 is violated if and only if a value vector can be produced which does not agree in any of its three places with the vector (1, 2, 3).

It is easily verified that there are eleven possible value vectors, including the four primary ones. Further operations of juxtaposition or crossing produce vectors already in the set. None of the eleven differs from (1, 2, 3) in all three places, so Theorem 10 is proved.

This new form of proof can readily be adapted to other formal systems and can suggest systems for which the theorem corresponding to Theorem 10 does *not* hold.

Like Varela's original proof, this one refers only to expressions with a single value re-entered (even though it may re-enter in more than one place). It is also possible to consider more elaborate forms of re-entry which are conveniently represented by systems of simultaneous equations. In another paper, for example, Varela‡ discusses the following system (in the course of proving his Theorem 5):

$$f_0 = \overline{f_0 f_1}$$
$$f_1 = \overline{\overline{f_1}\ \ \overline{f_1 f_0}} \qquad (1)$$

which can also be represented as

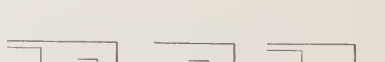
(2)

The treatment in terms of value vectors can be extended to deal with sets of simultaneous equations. Where the number of equations is n, each vector must contain 3^n elements, and the amount of computation required to list the possible vectors increases steeply with increasing n. For $n=2$ the number is 197 and it has been verified that there are no pairings of these which are not compatible with some assignment of values to the two overall expressions. The pair of equations (1), for example, is satisfied by any of the following pairs of values assigned to (f_0, f_1):

$$(null, \neg)\ \ (\square, null)\ \ (\square, \square)$$

Varela asserts that the r.h.s. of (2), and other similar expressions, cannot be reduced to *null*, $\square$ or $\neg$. Clearly, this assertion has to be interpreted with care; it certainly is not true to say that the r.h.s. of (2) is incompatible with any of these. However, it cannot reduce to one of them unless there exists some mechanism which coerces the inner loop, corresponding to f_0, to a compatible value.

It may be of interest to present a breakdown of the 197 value vectors possible when $n=2$, and a Theorem which can be inferred. There are two simple operations by which vectors can be transformed, namely by crossing, as already used in the $n=1$ case, and by interchange of the two inputs (corresponding to transposition if the vector is written as a 3×3 matrix). From any one vector, there are usually three others which are obtained by applying one or other or both of these operations, and it is only necessary to list 61 vectors to allow the derivation of all 197 by these transformations. Of the 61 vectors, 38 are members of groups of four, accounting for 152 of the total. One vector, namely that represented by ((2, 2, 2), (2, 2, 2), (2, 2, 2)), transforms into itself when either of these operations is applied. The remaining 22 vectors are members of groups of two and so contribute 44 to the total. All except one of the 22 are symmetrical in the two inserted values and are therefore unchanged by transposition. The remaining vector, which is ((2, 2, 1), (2, 2, 2), (3, 2, 2)) (or the transpose of this) is a member of a group of two because for it the operations of crossing and transposition are equivalent. This vector corresponds to the function



Since there is only one pair of possible vectors, apart from ((2, 2, 2), (2, 2, 2), (2, 2, 2)), with this property, it is possible to state the following:

‡F. Varela and J. Goguen, "The arithmetic of closure". In: *Progress in Cybernetics Research*, Vol. 3, edited by R. Trappl, G. Klir and L. Ricciardi. Hemisphere Publ. Co., Washington, 1977.

THEOREM *If a function* $h(p,q)$ *satisfies the following:*

$$h(p,q)=h(\overline{q,p})$$

then either $h(p,q)=\square$ *for all* p *and* q *or else*

$$h(p,q)=\overline{a\,|\,\square}\;\;\overline{b\,\square}\;\;\overline{a\,|\,b}$$

I have been unable to attach any deep signficance to this Theorem, but it has the intriguing property of producing a somewhat complex result from simple premises, and seems worth recording.

ALEX M. ANDREW
University of Reading
England

COMMENT ON "DYNAMIC AND LINGUISTIC MODES OF COMPLEX SYSTEMS BY HOWARD H. PATTEE

Sir—Molecular processes in the cell result in the transformation of information contained in a DNA molecule to active enzymatic processes. It is misleading to call the coded information in the DNA a description, to refer to the information as being in the linguistic mode, and as both of these imply to say that the descriptions are part of a language.[1] Language is generally considered to have the following features: (1) it is made up of units (words and symbols or if one prefers phonemes and morphemes) which have some systematic interrelation (2) the units (when taken in sufficient quantity to have meaning, e.g., words) have meaning by convention and (3) the system is used for communication. The second condition does not and it would appear cannot apply to molecular processes. The entire chain of processes involving DNA, m-RNA, t-RNA, amino acids, etc. proceeds according to the rules of biochemistry. A particular sequence of bases can lead to a corresponding chain of amino acids. There is no convention; there is a causal relationship. Such a correspondence is foreign to the nature of language. Genes do not describe, they contain information. Cells do not read messages, they transform information.

Pattee[1] in discussing the difference between the rate independent and the rate dependent modes concludes "the central problem may therefore be stated as the problem of complementarity". Although it is clear that he is referring to the interaction between "linguistic" and dynamic modes, for those familiar with the theory of quantum mechanics, complementarity implies a complex set of ideas. In extreme brevity, the notion of complementarity can be stated as follows[2]—classically: (1) natural phenomena obey exact laws—this leads to the principle of causality and (2) a physical theory should explain all phenomena as objects (or sets of vectors) existing in space and time; however, due to the fundamental wave-particle duality the above two requirements cannot be fulfilled. They represent complementary and mutually exclusive aspects of atomic phenomena. For the physical scientist the word complementarity has come to incorporate these notions by convention. In language, of course, words often have more than one meaning (whereas a particular triplet in the DNA code can only result in one amino acid). When Pattee refers to "dynamic and linguistic complementarity" there is no quantum mechanical dualistic problem. The masses of these systems are sufficiently large that observation is not limited in any significant way by the uncertainty principle. One could observe, in principle, the transformation processes which occur in a cell, discovering both the code and the structure at the same time.

REFERENCES

1. Howard H. Pattee, "Dynamic and Linguistic Modes of Complex Systems". *Int. J. General Systems* **3**, 77, pp. 259–266.
2. Werner Heisenberg, *The Physical Principles of the Quantum Theory*. trans. C. Eckart and F. C. Hoyt, Dover Publications, 1930. Max Jammer, *The Philosophy of Quantum Mechanics*. John Wiley, New York, 1974.

DENIS DONNELLY
Siena College
Loudonville, New York, U.S.A.

REPLY TO DONNELLY'S COMMENT ON "DYNAMIC AND LINGUISTIC MODES OF COMPLEX SYSTEMS", BY H. H. PATTEE

Donnelly's first criticism of my paper is that it is misleading to call DNA a linguistic description because real words have meaning only by con-

vention, whereas, he asserts, the interpretation of DNA proceeds only by "rules of biochemistry." The question, of course, is where these rules come from. One of the most striking generalities of molecular biology has been the realization of just how conventional all of these rules are. Monod, in his book, *Chance and Necessity*, has elevated these observations to the level of a "principle of gratuity" where the informational controls of the cell "have no chemical requirements to answer to." To quote Monod[6] further:

> Between the substrate of an allosteric enzyme and the ligands prompting or inhibiting its activity there exists no chemically necessary relationship of structure or of reactivity. The specificity of the interaction, in short, has nothing to do with the structure of the ligands; it is entirely due, instead, to that of the protein in the various states it is able to adopt, a structure in its turn freely, arbitrarily *dictated* by the structure of a gene.
>
> From this it results—and we come to our essential point—that so far as regulation through allosteric interactions is concerned, *everything is possible.*

Not only is Donnelly's assertion, "there is no convention" unsupported by evidence or informed opinion, but his either-or conclusion, 'there is only causal relationship" shows that he does not appreciate what complementarity is all about. There is, of course, *both* convention and causality in every complex system that I am discussing, and it is the necessity of complementary modes of description, the linguistic mode for the conventions or contraints, and the dynamic mode for the causes, that is the point of my paper.

Donnelly's second criticism is that my usage of complementarity is not, in his opinion, a part of the "complex set of ideas" which a physicist would associate with the word, and that the latter would not apply to coding events in the cell. Again, Donnelly offers no support for these statements. As I state in my paper, the complementary relation between description of linguistic rules or constraints and description of dynamical laws is closely associated with the symbol-matter (or subject–object) problem of epistemology and the measurement problem in physics. I did not go into the "complex set of ideas" that gave rise to complementarity because I believe it is still unclear even to expert physicists.[4,5,11] That is why I refer to the "problem of complementarity." However, in support of my usage, I offer the following arguments. While Bohr never gave a precise or formal definition of complementarity, he consistently used the term to apply to "modes of description" one of which derived from the idea of causality or the laws of nature, and the other derived from the idea of ordinary, intuitive space-time perception of events or observations. In classical physics these modes appear to work together, or they can be used separately, but in quantum mechanics they must be used together, and this gives rise to the problem.[2] If a physical system is used as a measuring device it must be describable both as an arbitrary constraint executing a rule (the linguistic mode) and as a causal system obeying the laws of motion (the dynamic mode). This situation, to quote Bohr,[1] "forces us to adopt a new mode of description designated as *complementary* in the sense that any given application of classical concepts precludes the simultaneous use of other classical concepts which in a different connection are equally necessary for the elucidation of phenomena." Bohr also saw this complementarity bearing "a deep-going analogy to the general difficulty in the formation of human ideas inherent in the distinction between subject and object." My specific association of what I call the linguistic mode with arbitrary, rate-independent, non-holonomic constraints, and the dynamic mode with inexorable, causal, rate-dependent laws is derived in large part from von Neumann's theory of measurement[10] in which he assumes two types of processes, the first kind, "the discontinuous, non-causal and instantaneously acting experiments or measurements" ("arbitrary changes by measurement"), and the second kind, "continuous and causal changes in the course of time," which follow the laws of motion. It is the main point of my paper that self-maintaining complex systems, (i.e., systems that do not spontaneously degenerate), require precisely this type of complementary description for which neither a rate-independent model of its linguistic constraints nor a rate-dependent model of its dynamics provides and adequate representation of the system's behavior.

Donnelly concludes with another unsupported assertion that, "one could observe, in principle, the transformation processes which occur in a cell, discovering both the code and the structure at the same time." If, as I have proposed,[7,8,9] the reading of genetic information requires measure-

ient by the coding enzymes (the synthetases and RNAs), then the measurement of the electronic tructure of the enzymes will necessarily interfere vith the dynamic execution of its rate-dependent unction, [cf. Bohr[3]]. This is one argument why omplete self-description of any system is a physi-al impossibility, though not a logical or formal mpossibility.

However, I would not like this specific point to bscure my more general argument that general ystems theories cannot be expected to provide dequate models of biological, social or political ystems, which obviously function through their wn internal descriptions, until the epistemologi-al problem of complementarity between subject nd object is more clearly recognized. My argu-ients are summarized in the Conclusion of my aper, and I do not believe they are strongly ensitive to various interpretations of quantum iechanics.

REFERENCES

1. N. Bohr, *Atomic Theory and the Description of Nature.* Cambridge University Press, 1934, p. 10.
2. N. Bohr, Biology and atomic physics. In: *Atomic Physics and Human Knowledge*, John Wiley, New York, 1958, pp. 20–21.
3. N. Bohr, Quantum physics and philosophy—causality and complementarity. In: *Philosophy in the Mid-Century*, R. Klibansky, ed., La Nuivo Italia Editrice, Florence, 1958. Reprinted in *Essays 1958–1962 on Atomic Physics and Human Knowledge*, Interscience, New York, 1963.
4. D. Bohm, On Bohr's views concerning quantum theory. In: *Quantum Theory and Beyond*, T. Bastin, ed., Cambridge University Press, 1971.
5. M. Jammer, *The Philosophy of Quantum Mechanics.* John Wiley, New York, 1974, Chapter 4.
6. J. Monod, *Chance and Necessity.* Alfred Knopf, New York, 1971, pp. 76–78.
7. H. H. Pattee, The physical basis of coding and reliability in biological evolution. In: *Towards a Theoretical Biology.* Vol. I, C. H. Waddington, ed., Edinburgh University Press, 1968, pp. 67–94.
8. H. H. Pattee, Can life explain quantum mechanics. In: *Quantum Theory and Beyond.* T. Bastin, ed., Cambridge University Press, 1971.
9. H. H. Pattee, Physical problems of decision-making constraints, *Inter. J. Neurosciences* **3**, pp. 99–106, 1972.
10. J. von Neumann, *The Mathematical Foundations of Quantum Mechanics.* Princeton University Press, 1955, p. 417.
11. C. F. von Weizsäcker, The Copenhagen interpretation. In *Quantum Theory and Beyond*, T. Bastin, ed., Cambridge University Press, 1971.

SIMULATION

The leading journal devoted to the methods and applications of dynamic modeling and simulation using all types of computers.

Contents *VOLUME 31: NUMBER 3: SEPTEMBER 1978*

Calendar of short courses . . . . . *ii*

Calendar of events . . . . . *iv*

Presenting . . . Richard Bronson . . . Norman F. Schneidewind . . . . *vi*

ARTICLES

Large-scale simulation of brain cortices 73
Larry D. Wittie

Evaluation of startup policies in simulation experiments 79
James R. Wilson A. Alan B. Pritsker

FRONTIERS IN SYSTEMS MODELLING

Simulating the growth of cellular forms 90
Pauline Hogeweg

SENSIBLE SIMULATION FOR INDUSTRIAL PROCESSES

Simulation of conveyor systems with dual inputs 97
E.A. Elsayed H.A. Elayat A.M. Ragab

TECHNICAL COMMENTS

Mathematical modeling made easy (for fun and profit) 105
Dennis W. Baker

Simulation: a saturated art limited by human capabilities 107
L. Dekker

Call for papers . . . Classified advertising 108

SIMULATION IN THE SERVICE OF SOCIETY

Can large-scale simulations alleviate world resource problems? . . . *vii*

Membership Application . . . . . *xi*

THE SOCIETY FOR COMPUTER SIMULATION
(Simulation Councils, Inc.) P.O. Box 2228, La Jolla, California 92038

nt. J. General Systems
979, Vol. 5, pp. 63–71

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

Special Book Reviews

The editor invites books for review on any subject relevant to general systems. Books for review must be sent to Dr. George Klir, Editor, International Journal of General Systems, *School of Advanced Technology, State University of New York, Binghamton, New York 13901, U.S.A. (and* not *to the publisher).*

n a recent paper (Self-Organization of Living ystems: A Formal Model of Autopoiesis, *Int. J. General Systems*, 6, 13–28, 1977) I have referred o the works of B. Trentowski, A. A. Bogdanov, . Leduc and J. Ch. Smuts. Although I had been uoting all four authors for many years, there vas an unexpected flurry of inquiries this time, s to the details and specific whereabouts of the nentioned publications.

Because two of them have never been trans-ated into English (originals in Polish and Russian), one was last translated in 1911 (original n French) and the one in English was written by Boer general, most of these books might not be eadily accessible to the average reader. That is, f course, a great pity.

After Professor Klir's expression of interest in dusting off" some of the "paradigms lost", which hese books undoubtedly represent, we have de-ided to prepare their reviews and present them o the readers of *General Systems* as a self-ontained package. I asked one of my students, Velimir Srića, to assist me in preparing the notes, inalizing the translations and formulating new nterpretations. His enthusiasm and insight show hrough the reviews. His attitude seems to exem-lify and confirm my intuitive feeling that these our authors and their respective messages are going to appeal especially to the youngest gene-ation of researchers.

We hope that we shall ultimately find a pub-isher who would undertake the task of translat-ng and re-issuing this series and thus help to einstate some of the lost paradigms to their proper place in the history of scientific thought.

MILAN ZELENY
Columbia University, New York, U.S.A.

STOSUNEK FILOZOFII DO CYBERNETYKI CZYLI SZTUKI RZADZENIA NARODEM

(On the Relationship Between Philosophy and Cybernetyka: the Art of Managing a Nation) by Bronislaw Trentowski, published by J. K. Żupański, Poznań, 1843, 194 pages.

Norbert Wiener, the founder of the science of cybernetics, has the following to say about the coinage of the term cybernetics in *The Human Use of Human Beings*:

> Incidentally, I found later that the word [cybernetics] had already been used by Ampère with reference to political science, and had been introduced in another context by a Polish scientist, both uses dating from the earlier part of the nineteenth century.

Let us search out the Polish scientist and explore the "another context" of the art of steersmanship, the ancient kybernetike of Plato's "Republic". Let us also remember that Ampère's usage of cybernétique, the science of men who govern, dates from about the same time as Trentowski's treatise of 1843.

We shall not engage ourselves in discussing the term itself and we shall withdraw our judgment with regard to its primary or proper usage. We need the essence, the distillate of Trentowski's major theses, as we attempt to reinstate explicit conflict between alternative theories, that unfailing sign of scientific vitality. Trentowski derived Cybernetyka from the Greek word $\chi\upsilon\beta\varepsilon\rho\nu\acute{\alpha}\omega$, or "guberno", i.e., a political or governmental district. Cybernetyka then is the art and science of human systems management, the art of governing complex human organizations in their various manifestations as political, economic, cultural and other institutions.

The practitioner of this art (Trentowski calls him *Cyberneta*, the steersman) is gubernator, the one who manages complex human systems. To practice the art of management without any closer study of the underlying theory is quite similar to the practice of a common healer who presumes to treat others without deeper under-

standing of medical science and without mastering any of the art of medicine.

Trentowski insists that human component should be the main object of managerial concerns:

> Gubernator [or Cyberneta] is dealing with political history, radicalism and philosophy at the same time. Not through the newspapers, journals, books or libraries, but through the *people*, with all of their *real, living and strong* partialities. To reconcile their different views and biases he cannot and should not build a new political system, as would a philosopher attempt to do, because *humans are neither mathematical symbols nor logical categories.*

Trentowski thus sets Cybernetyka apart from certain trends in mathematical modeling. He foresaw in 1843 that our ability to mathematically allegorize complex human organizations and processes is not only limited but the attempts to ignore such limits might not be appropriate. In contrast with modern critics of cybernetics, he does offer an alternative:

> In summary, the gubernator does not design the future, as would any radical philosopher try; rather he lets the future to be born on its own, independently; he only offers assistance as an experienced and skilled *social obstetrician.*

Complex human system, its organization and its structure cannot be designed modeled or implemented in analogy with engineering systems, mechanical contrivances and man-imposed hierarchies. They are to be managed through all the biases and partialities of human action, decision making, intuition and imagination.

Trentowski clearly recognized the presence of self-organizing, fuzzy, dynamic, interactive and organic attributes in complex human organizations. These are essentially irreducible to the mechanism of a first-order, negative feedback modality. Loops and double-loops of modern organization theory seem to be pathetically inadequate when compared with his simple insight:

> I reiterate once more: it is not that the gubernator develops political reality, but that the humanity itself keeps evolving under his steersmanship. It is not him but human reality who invents and implements its own political theories. To master such self-organizing process requires quite different concept of management; it requires wisdom far above that of a philosopher.

The notion that static political or philosophical models could either replace or substitute for Cybernetyka are also sincerely doubted:

> The gubernator does not find much of importance within the general philosophical-political theories. Under the same theoretical ideology he would govern quite differently i England, differently in France and Spain, differently i Austria, Russia or Prussia. Also, in the same country, h would govern differently today or tomorrow.

The gubernator does not presume to unrave the secrets of humanity but assists the humanit in proceeding on its own self-evolutionary path His mode of inquiry is not disciplinary, cyber netyka is not only a science. It is a trans disciplinary synthetic art relying only partially o the main results of individual scientifi disciplines.

Trentowski concludes with an interesting, a though quite flowery, allegory on cybernetyka role and importance:

> History, that's an old house, either standing status quo or ju lying about and gradually regressing into oblivion. Or, at th best, receiving a few, mostly insignificant repairs here an there. Radicalism is like the continual demolition of the hous blowing up its helpless adorations with a gunpowde Philosophy represents the architect's plan for a new hou that would be more suitable to the present circumstance What then the role of Cybernetyka should be?
>
> A philosopher would advise: "Let it be the building of th new house, according to my architectural drawings." But wa a minute, brother! Is it only you who is entitled to ind pendent thought? Is the gubernator to be just a thoughtle performer of your plans!? You want to be the master of th social house and make the gubernator your bricklayer! I, to have my own ideas and they are even more important tha yours because I possess the final authority. I am the actu administrator of the state and I am generally recognized such. I steer the ship of the present and I take full r sponsibility for my actions, whether or not the historical glo ensues. Thus, I allow nobody to keep wrestling for the rudd and interfere with the course. I am the master of my field much as a philosopher is the master of his.

Thus, the manager is "the one at the helm", th steersman and the catalyst, not a designer, organ izer, performer or a controller. Managemen then is an art rather as much as science. It is no the "naive" or "primitive" art of untrained, exper mental laymen, regardless what the market valu of their "art" happens to be.

The true value of Trentowski's Cybernetyka i in that human beings, human action and bot the fuzziness and the complexity of human sys tems are built in its very foundations. It unlikely that building upon such foundation would ever lead to our current abuse of the firs order, negative feedback scheme.

It is not a unitary science which applies a singl simple behavioral mechanism to the most con plex systems or processes in a variety of differe fields, including biology, psychology, etholog

nguistics and economics. An observer should ot impress the unity of a new conception on uch fundamentally different systems. Unitary inds, unitary paradigms and methodologies will ot unify the observed world, only our per-eption of it.

There are several ways of interpreting the neaning of "unitary science". It could be a cience providing a single, unifying point of view, single explanatory paradigm for all other sci-nces. All of its adherents would hold the same iew and attempt to spread its gospel across the isciplinary boundaries.

Or, it could provide a framework or a matrix hrough which the individual sciences could join nd interact with regard to a given, common urpose. The variety of views and paradigms vould be preserved and even enhanced. Such Jnity through Diversity would be natural, un-mposed and the creative interdisciplinary tension vould be restored.

TEKTOLOGIA: VSEOBSHCHAYA ORGAN-ZATSIONNAYA NAUKA (Tectology: Universal Science of Organization) by A. Bogdanov, published in 1922 by Z. I. Grschebin Verlag, Berlin, Volumes 1, II and III, 527 pages.

A system is organized not in general but only in relation to ome definite actions, reactions, energies; it can be at the ame time disorganized in relation to some activities, and eutral with respect to others. (p. 80)

Organization can be defined as the state in which the ctivities are combined more successfully than the resistances pposing these actions. (p. 71)

Highly organized system is the one that counteracts nu-nerous activities in its environment. (p. 81)

Only growth of activities of a system at the expense of its nvironment insures the preservation of that system. (p. 160)

These randomly selected quotations are not rom some modern book on GST but from "Tectology" (Universal Science of Organization) vritten in Russia by A. Bogdanov in about 1912. The author (1873–1928) was a medical resear-her, philosopher, writer, economist and poli-ician, and although L. von Bertalanffy and other uthoritative contributors to the systems science paradigm have no reference to Bogdanov's monu-nental work, it may be said that it is histori-ally the first expanded variant, less mechanistic nd more profound, of what is today called GST General Systems Theory).

Tectology (Greek word "tekton" means "buil-der") is a theory about creation, maintenance and destruction of systems viewed as "organized complexes".

In nature and society there exists certain unity of organizational methods that can be studied scientifically. (p. 29)

Organizational processes are universal in the small and in the great, in the living and in the dead. (p. 527)

The primary elements of tectology are activities, resistances and their combinations; "organized activities could be described in terms of resistances that they overcome" (p. 68). Since every activity is a resistance to other activities with which it interacts, there are three possible situations: the organized whole which is more than the sum of its parts (synergy), the disorganized whole which is less than the sum of its parts (disynergy), and the neutral whole which is exactly the sum of its parts. The latter one, a neutral complex in equilibrium, represents a static case to which mathematics (or tectology of neutral complexes) can be applied. The task of tectology is to deal with the dynamics of organizations and with the subtlety of moving equilibria, resulting from more or less successfully combined activities, opposed to their resistances. In order to accomplish that, tectology "must not only develop its methods but also investigate and clarify them; because of this tectology represents the completion of the circle of sciences" (p. 83). Developing the appropriate symbolism is therefore the first and the most complex task, since tectology "views phenomena in their full completeness and complexity" (p. 83).

Based on that, Bogdanov describes formulation of basic organizational mechanisms that cause creation, regulation and destruction of systems. "Conjunction" occurs when complexes interact together. This can result in different stages of cooperation among the activities of the complexes involved. If the activities "fit" together, the common or intermediate links are created. If the activities mutually "paralyze" one another, the "tectological boundary" separating the complexes emerges. The first situation is called "conjunctive crisis" (crisis C), and the second "disjunctive crisis" (crisis D). Both crises are really manifested in a breach of the existing systemic equilibrium, and both constitute an organizational transition to a new equilibrium. Bogdanov's point is that, by studying tendencies of these organizational crises we may be able to predict the final outcomes or the limiting equilibria toward which a system tends.

For a system, the environment represents a universe of antagonistic forces in which assimilation and disassimilation of its requisite variety permanently take place.

> Only growth in activities, preponderance of assimilation, that is growth of activities of a complex at the expense of its environment, insures the preservation of that complex. (p. 160)

The case where assimilation is greater than disassimilation, Bogdanov denotes as "positive progressive selection", while the opposite situation is called "negative progressive selection". The positive selection results in increased complexity and heterogeneity of internal relationships through supply of resources from the environment—progressive selection which is negative, causes a removal of all the volatile, discordant and antagonistic in a system, resulting in increased homogeneity, coordination, order and systematization. These two activities, serving as a base of systems regulating mechanism, have an impact on stability of organizational forms. Positive selection adds new contradictions, thus weakening the whole from a structural viewpoint, while negative selection removes the weakest points and thus tends to increase structural stability at the cost of lowering quantitative stability.

The process of concentrating the already gathered organizational content is enhanced by a "centralist", a complex that exhibits prepondering influence on other systems (a leader in a group, for example). The more powerful a centralist, the more "plastic" a system (plasticity is defined by Bogdanov as the relative ease of forming new connections and regrouping of systemic elements). System plasticity is characterized by the richer and greater adaptability to environmental changes, but also by greater vulnerability of thus increasingly complex whole. Just the opposite characteristics could be found with "skeletal" complexes that are durable and less organized, therefore relatively fixed. Both aspects of a system help and support balanced struggle of inner creation and inner inertia. Bogdanov gives an example: words and symbols, as the skeletal part of human communication, tend to fix, strengthen and protect the living plasticity of mental images against decomposition.

In favorable environments the plastic part, assimilating the requisite variety, tends to grow faster and is constrained by the skeletal part. When the environment is unfavorable, the skeletal part protects the plastic part of disassimi lation and thus ensures survival of the whole.

A high degree of correspondence to a give environment means a lack of such correspon dence to different environments. Therefore, th more conservative the environment, the longe selection processes continue in the same direc tion. For positive selection to continue, totality o favorable conditions is needed; for negative selec tion to take place, a single unfavorable conditio suffices. Therefore, the second law of thermo dynamics is the loser in a long-run; the phoeni bird of creation is inexhaustible, while "negativ selection is continuously interrupted by its ow exhaustion" (p. 445).

A change in organizational form of a comple is called crisis (Greek "krisis" means "solution" "turning point"). This is a disturbance in a equilibrium as well as a transition to a new one.

Organizational dialectics describing an organi zational act also uncovers the ubiquity of it phases: "uncommitted conjunction" of elements i followed by "systemic differentiation", and th derivatives of this disjunction participate in "systemic consolidation". "Any continuity ca be broken down by analysis into an infinite chai of crises" (p. 497), therefore the notion of crisis i universal. The same is true for organizationa processes.

"All the time the Guard was looking for he first through a telescope, then through a micro scope, and then through an opera-glass. At last h said: 'You are travelling the wrong way'" (Lewi Carrol): These words may be appropriately ad dressed to the general systems theoreticians an to the proponents of cybernetics in social science who look at the world through the glasses of sti prevailing paradigms that cannot answer most o the "soft" questions regarding so called "living systems, especially those having humans as thei elements.

From a viewpoint of the new "human system management" paradigm, A. Bogdanov was, al though very much alone, travelling the right wa He clearly understood the need for a transdiscip linary approach in dealing with organizationa experiences in different fields. His holistic view o the Universe was deep and profound.

A system (or a complex; these two word Bogdanov uses interchangeably) is defined as a interaction and interrelationship of elements suc that the parts are as much the result of the whol as the whole is the result of the part

Furthermore, any complex is contained in its environment simultaneously as a casting material and a formulating model: "the more similar is the organizational material and conditions influencing this material, the greater similarity may be expected in the appearing organizational forms" (p. 460).

In the process of organizational design, humans have a role of observers, as well as participants; they are not able (unlike Nature in the case of positive selection) to deal with all, but only with some aspects of an object of their perception. Since men select representatively, they are more successful in destructive activities (for negative selection to occur only one unfavorable condition is enough). Furthermore, observers cannot see the whole because they are within one.

Bogdanov also clearly understood that interactions among components (the "organization") are what makes a system distinct, and not the properties of components themselves. The elements of a system are bound together by a unity of purpose, performed functions and the environment. The "structure" that shows to an observer obscures the reality the same way as one gets deceived by a static illusion of two water currents going in the opposite direction, a dissipative structure.

The theory of self-organization of complex systems can therefore find its precursor in Bogdanov, who actually first described the "autopoiesis" of a political party (p. 502–504), or the role of "catalyst" ("eggressive center", "centralist") in self-creation of an organization. His refreshing and stimulating book also contains the basic elements of a general theory of crises (or catastrophes, if you wish).

Bogdanov added a significant contribution to the dialectical analysis of social systems, elaborating the idea that more than a simple possession of the means of production is needed, in order to exercise power over a society. What he considered more important was knowledge of how to organize social activities in a productive manner (the Soviet practice in agriculture strongly supports that notion).

The writer of "tectology" was also a proponent of spontaneous generation of "life", as he quoted, unfortunately nowadays long forgotten, experiments in the field of synthetic biology performed by Leduc, Bütschli, Quinke *et al.* (artificial cells, osmotic growths) at the turn of the century.

If a man is to be judged by his ideas, we must have nothing but admiration for A. Bogdanov. Among all the books that were suppressed or forgotten, his "Tectology" is maybe the one that is really worth rediscovering.

HOLISM AND EVOLUTION by J. Ch. Smuts, Greenwood Press, Publishers, Westport, Conn., 1973, 250 pages. (Originally published in 1926 by Macmillan and Co., Limited, London).

Alfred Tennyson wrote a poem about a flower growing in a stone fence. After extricating it and almost cutting it into pieces, analyzing it as if it was a set of autonomously significant parts, he meditated and thought he would understand World, God and Nature if he could descend into the basic laws of the life and growth of the flower. Mazuo Basho, Japanese haiku poet, in dealing with the same theme, was trying to penetrate a flower mentally without even touching it. He was puzzled by the wholeness of Nature and her intriguing complexity and he tried to gain an understanding of Life from within the flower, not from without.

The juxtaposition of these two approaches emphasizes not only the major difference in spiritual and scientific attitudes characterizing Western and Eastern thinking, but shows also the beauty of wholes and the exactness of parts.

General Jan Christiaan Smuts (1870–1950; a founder and Prime Minister of the Union of South Africa from 1919–24, and also one of the founders of the United Nations in 1939–48) could not have agreed more with Basho. His Holism and Evolution", first published in 1926, represents an attempt to reunify the analytically shattered scientific parts into a holistic unity. Due to its prematurity and other circumstances, it became a "paradigm lost", one of those grayish books from grandfather's library, sunk into dust and oblivion.

In his study, General Smuts tried to explain the concept of Holism, the self-creation of wholes, viewed as the *reflection of an inner character of the Universe.* In order to do so, he deals with the "debateable borderland between science and philosophy", reexamining the notions of Matter, Life and Mind:

> Matter, life and mind are, so to speak, the original alphabet of knowledge, the original nuclei of the experience, thought and speculation. (p. 6)

Narrow mechanical and cause-effect conceptual systems, according to Smuts, do not apply to reality and "make it inexplicable in terms of relations between things, qualities and ideas" (p. 17). Holism is a fundamental principle of which matter, life and mind are the progressive outcomes. Holistic unity of elements is always more than the sum of the parts:

The synthesis affects and determines the parts so that they function towards the "whole" and the whole and the parts therefore reciprocally influence and determine each other, and *appear more or less to merge their individual* characters: the whole is in the parts and the parts are in the whole, and this synthesis of whole and parts is reflected in the holistic character of the function of the parts as well as of the whole. (p. 86)

Holism is creative and self-creative; final structures are more holistic than the initial ones (three fundamental structures of the universe are the atom, the cell and the mind): There is a progressive gradation of holistic synthesis in nature from mere physical mixtures, chemical compounds, organisms, "minds" to Personality, which is the most evolved whole among the structures of the Universe. The philosophical "esse" is therefore neither idealistic nor materialistic, rather it is holistic.

Ultimate reality of the universe is neither matter nor spirit, but wholes as defined in this book (p. 117) a balanced correlation of organs (parts) and their functions such that *through the holistic action components become unrecognizable in the unified result*).

The gradual creation or self-creation of wholes takes place through the progressive series of *structures* from the beginning to the end, through the process of Evolution. Simply stated, evolution represents a trend from mechanical toward holistic through the "intra-penetration" of simpler structures (parts), merging together as a result of the inner holism of the universe.

The structures of matter, mind and life are on quite different levels of *organization* and inwardness, the one acts inside the other and through the other. (p. 268)

Holism is not Monadism; spirit (or soul) is not attributed to atoms and chemical structures, but it arises as a result of holistic evolution in such a way that the world of matter, life and mind can be grouped into a series of *whole-creative processes*. Thus "omne vivum e vivo" is substituted by "all wholes from wholes".

The conception of wholes covers a much wider field than that of life, that its beginnings are traceable already in the organic order of nature, and that beyond the ordinary domain o biology it applies in a sense to human associations like th state and other human activities and creations of huma spirit. (p. 98)

General Smuts was more a man of intuitio than a scientist, therefore his book is an in exhaustible source of interesting ideas and no tions, as well as doubts, indications or unanswer able questions. Above all, it is a great book o systems (if we agree with C. W. Churchman tha the essence of the systems approach is confusio as well as enlightenment, we will be bound t accept and appreciate both these aspects i Smuts' Holism). Moreover, it could become "paradigm regained", the old bottle for the ne wine. Let Mr. Smuts speak for himself:

An organism is a *continuous autogenesis*. It can only b explained by reference to its past and its future as well as it present; the central *structure* is not sufficient and literally ha not enough in it to go round in the way of explanatio (p. 79)

Mere preoccupation with detailed mechanisms will n longer suit the immensely enlarged scope of present-da science. (p. 317)

The holistic transformation of causality takes place in a organic stimuli and responses. The cause or stimulus applie does not issue in its own passive effect, but in an *activ response which seems more clearly traceable to the organis or whole itself.* (p. 119)

Compared to its parts, the whole constituted by them something quite different, something creatively new; creativ evolution synthesizes from the parts a new entity not on different from them but quite transcending them. *Whole always transcendent to its parts and its character cannot b inferred from the characters of its parts.* (p. 128)

Behind the evolutionary movement and the holistic "field of Nature, is the inner shaping, directive activity of Holis itself, working through the wholes and in the variations whic creatively arise from them. These variations are controlled regulated expression of the *inner holistic development* o organisms as wholes. There is direction and control righ through the entire forward movement, not only in the origi of variations but also at the various subsequent stages of thei internal and external "selection". (p. 342–343)

Let us now very briefly summarize some of th basic ideas of the new, emerging paradigm o systems science, the "systems autopoiesis":

—all living systems are structurally coupled wit their environment, therefore inseparable fro it:

—organic systems cannot be explained throug their structure, or through the properties o their components, but only through the ne work of interactions among components (o organization);

-organic systems are characterized by spontaneous order; the components of an autopoietic system through a network of interrelated component-producing processes generate again an autopoietic system;

-autopoietic systems are closed systems with no apparent inputs or outputs; therefore, the stimulus-response framework should be substituted by the notion of state-dependency;

-autopoiesis of organic systems is opposed to the mechanistic "cybernation approach" to systems design through a series of feedback loops.

The compatibility of these two frameworks is uite impressive. Smuts' intuition has already een supported by some researches using comuter simulation or modern laboratory quipment.

It is worth noting that in "Holism and volution" Smuts gives also one of the clearest escriptions of what has recently become known s a *theory of fuzzy sets*:

Ve have to return to the fluidity and plasticity of nature and xperience in order to find the concepts of reality. When we o this we find that around every luminous point in exerience there is a gradual shading off into haziness and bscurity. A "concept" is not merely its clear luminous centre, ut embraces a surrounding sphere of meaning or influence of naller or larger dimensions, in which the luminosity tails off nd grows fainter until it disappears. The hard and abrupt ontours make reality inexplicable, not only in the case of ausation, but in all cases of relations between things, quaties and ideas. (p. 17)

As most of the books in which the author tries o convince his readers about something he is not eally sure of, but strongly feels it is true, Holism and Evolution" is redundant and somehat unclear. If there are two kinds of eople: those with tradition-shattering attitudes, nd those with tradition-bound attitudes, then muts' book is most likely not to be recomnended to the latter group. It would, however, e very much appreciated by those who believe, s Plinius did, that there is really never too much f what there should be most.

THE MECHANISM OF LIFE by Stéphane educ, Rebman Limited, London, 1911, 172 ages.

cience shares some common features with sucessful politics: the leaders should sound convincng and self-confident, the ideas should appeal to ne majority of listeners, satisfactory answers should be given to the burning questions. At the least, in both spheres the rules are dictated by the winners (so called *normal science* is often defined as the formative matrix of beliefs espoused by a given scientific community at a given time. It often suppresses fundamental innovations because they might introduce conflict with regard to its basic commitments).

Professor of medicine Stéphane Leduc was one of the losers. In 1907 the French Academie of Sciences excluded from its "Comptes Rendus" his report on some experimental researches on diffusion and osmosis because it expressed notions of a theory of life based on spontaneous generation. When he published his "Mechanism of Life", he had already been expelled from the scientific community.

Although recalling the losers might seem quite strange in present-day scientific and cultural setting, let us quote some of the main points Leduc had made in 1911, in his today extremely scarce book.

The living organisms are transformers of energy, matter and form and these metamorphoses are necessarily in certain *harmony* which we call life. All the phenomena of nature are *continuous in time and space*: "if we find it impossible to discover the line which separates life from other phenomena of Nature it is in fact because no such line of demarcation exists—the passage from animate to inanimate is gradual and insensible." (p. xiii).

The higher levels of organization of matter are a result of *spontaneous generation based on inherent physico-chemical forces* that are "quite sufficient to produce forms like those of living beings—cavities—liquids—tissues—membranes." (p. 153).

Since the liquids are the most important constituents of any living organism, Leduc concludes that "the junction of two liquids of different concentration is the arena in which takes place both the chemical transformation of matter and the correlative transformation of energy." (p. 36). Therefore, in explaining life, *diffusion and osmosis* are the crucial factors.

An excellent experimentator Prof. Leduc, following and integrating the findings of a large group of synthetic biologists (A. Nollet, M. Traube, E. Montgomery, Harting, D. Monnier, K. Vogt, Quinke, G. Rose, Link, Runge, R. Böttger, Reinke, F. Cohn, H. de Vries, Tamman, Graham, Bütschli, Famintzin, Hansen, M.

Benedikt, Dubois, Herrera *et al.*, pp. 115–122), dedicated an enormous amount of time, effort and energy to the exploration and research of interactions among different solutions. He accumulated a large collection of life-like shapes and structures, *self-created from inorganic matter*, cells of a solute surrounded by osmotic membranes. One of the many typical experiments is described as follows:

> On a glass plate we spread 5 per cent. solution of pure gelatine and when set sow on it a number of drops of a 5 to 10 per cent. solution of ferrocyanide of potassium. The drop must be placed at regular intervals of 5 mm. all over the plate. When these have been allowed to diffuse and the gelatine has dried, we obtain a preparation which exactly resembles the section of a vegetable cellular tissue. (p. 62)

In a similar fashion Leduc obtained so called *osmotic growths* of different solutes yielding such remarkable shapes and forms as those of flowers, fungi, shells, leaves.

> By varying the solution or concentration we may thus grow an infinite variety of *shell forms*. (p. 136)
>
> The resemblance is so perfect that some of our productions have been taken for *fungi* even by experts. (p. 134)
>
> We may also by appropriate means produce terminal organs resembling *flowers and seed capsules*. (p. 131)
>
> The foliaceous expansions, or osmotic leaves, also present great varieties both of appearance and of structures. (p. 143)

Thus created osmotic growths feed on their environment and excrete their wastes. They grow old, they die when they "have expended the osmotic force contained in the interior of their capsule" (p. 141). They could be rejuvenated and certain of these artificial conglomerates of cells may even be made to grow out of the solution into the air (so called "aquatic" and "aereal" cells exhibit specialization in form). They incorporate also the phenomena of circulation and respiration and the process of nutrition. An injury to an osmotic growth can be repaired by coagulation. Some of the products are capable of performing periodic movements. They also imitate the colors, the forms, the shapes as well as the texture or the microscopical structure of organic growths.

A new field in biology emerges: *morphogeny*—synthesis of living forms; *physiogeny*—synthesis of functions; and *synthetic organic chemistry*. These three represent the branches of *synthetic biology*.

In explaining his experimental findings, Ledu offers several interesting points:

—a third law of thermodynamics, *the law o stimulus*, explains the growing complexity o living systems: "a living being exhibits tw sorts of movements, those which are the resul of stimulus from without, and those which ar determined by an excitation arising fron within. In the higher animals the stimulus o exciting energy coming from the entourag may be infinitely small when compared wit the amount of energy transformed. Moreove the response to an identical excitation may s vary as to give to these different responses a appearance of spontaneity." (p. 155); and, "sen sations represent the intervention of an exte nal exciting energy (catalyst, ferment, enzyme which provokes the response." (p. 112)

—no sharp division between inanimate natur and life exists, therefore life is a matter o *harmonious organization of substances involve* rather than a function of the properties of th substances themselves. "It does not seem tha the classifications depending on differences o substance are any more important than thos which depend on difference of form." (p. 159)

—*Lamarckian ideas on evolution and spontaneou generation* are adopted in explaining life an its self-creation, self-organization or sel reproduction. "The theory of evolution nec essarily admits the physico-chemical nature o life and the fact of spontaneous generation (p. xv); "the ordinary physical forces have i fact a power of organization infinitely large than has been hitherto supposed by the boldes imagination" (p. 167)

—*periodicity* is inherent in all the processes tha involve life; it is reflected in an interna "biological clock" as a feature of sel organization of living systems. 'All the phenc mena of life are periodic. The movement o heart and lungs, sleep and waking, all nervou phenomena have a regular periodicity. It i possible that a study of these purely physica phenomena of periodic precipitation may giv us the key to the causation of rhythm an periodicity in living beings." (p. 75)

It seems extremely paradoxical that such a organic book should include the term "mech anism" in its title. It seems even more para

oxical, that the exceptionally promising field of ynthetic biology should have waited so long and s still waiting to be rediscovered. If one has oubts about the experimental results Leduc ob-ained (they are accompanied by a large number f photographs and detailed descriptions of con-entrations, external conditions as well as other roperties of participant chemicals), then one nay take a few elementary courses in chemistry nd easily repeat most of the basic experiments. Even that would be less involved than duplicat-ng the complicated "experiments" by which pre-ent biologists try to "create" protein-like com-pounds by bombarding different chemical soups vith enormous amounts of electricity. Even when hey succeed, the existence of scattered amino-cids in that soup proves nothing. The parts lone can never properly explain the whole. As Leduc states, "if albumin is ever produced by ynthesis in the laboratory it will probably be lead albumin." (p. 157).

Not only that a rediscovery of Leduc's book nay mean a new page in *modern biology*, but its chapters have proven a valuable source of notions, ideas, concepts and supports for the new paradigm of "*autopoietic organization of complex systems*": "living" systems that are generated by a spontaneous order, involving catalysts as internal activators and modificators of "life", with state-dependent behavior characterized by the properties of interactions among components and not by the properties of components themselves. *Paleontologists* facing Leduc's shells and plants identical in shape with those of organic origin, might be encouraged to reconsider some of their classical conclusions and findings. *Philosophers and poets* might find some new sources of contemplation regarding the painful and intuitive recognition of life and absence of supernatural.

As a high school student, one might still remember the way we were all laughing at the ridiculous idea of spontaneous generation ("worms emerging from mud"). After reading Leduc's "The Mechanisms of Life" one might be willing to feel the way Tertullianus had expressed his religious attitudes: "Credo quia absurdum".

TIMS MANAGEMENT SCIENCE

Volume 23, Number 11 July 1977

The Substitution of Communication for Transportation—A Case Study . . . . . David A. Lopez & Paul Gray 1149

A Model for Planning the Transition to Equilibrium of a University Budget . . . . . David S. P. Hopkins & William F. Massy 1161

How Often Will the Firemen Get Their Sleep? . . . . . John D. Finnerty 1169

An Evaluation of Flow Shop Sequencing Heuristics . . David G. Dannenbring 1174

An Algorithm for Deriving the Capital Market Line . . Gordon J. Alexander 1183

Multilocation Plant Sizing and Timing . . Ram C. Rao & David P. Rutenberg 1187

Optimizing Claims Fluctuation Reserves . . . . . Christoph Haehling von Lanzenauer & Don D. Wright 1199

Geometric Approaches to Solving the Traveling Salesman Problem . . . . . John P. Norback & Robert F. Love 1208

Incorporating Thresholds of Indifference in Probabilistic Choice Models . . . . . K. S. Krishnan 1224

An Investment Strategy with Overshoot Rebates Which Minimizes the Time to Attain a Specified Goal . . . . . Donald C. Aucamp 1234

NOTES

Capital Budgeting of Interdependent Projects: Activity Analysis Approach . . . . . Juval Goldwerger & Jacob Paroush 1242

A Comment on Rao's Class $\mathfrak{M}$ and $\bar{S}$ Matrices . . . . Gary J. Koehler 1247

A Note on the Dynamic Inventory Problem with Unknown Demand Distribution . . . . . S. Papachristos 1248

A Computational Refinement for Discrete-Valued Dynamic Programs with Convex Functions . . . . . Don R. Robinson 1251

COMMUNICATIONS TO THE EDITOR

A Comprehensive Model for Long Range Academic Strategies . Paul Gray 1256

Response to "A Comprehensive Model for Long Range Academic Strategies" . . . . . Anthony S. Walters & Janet Mangold Burnstein 1257

Notes about Authors 1258

Reprints of individual articles are not available; single issues may be purchased at $4.50 per copy.

Int. J. General Systems
1979, Vol. 5, pp. 73–74

© Gordon and Breach Science Publishers Ltd.
Printed in Great Britain

Calendar

January 3–8, 1979
TWENTY-THIRD NORTH AMERICAN ANNUAL MEETING OF THE SOCIETY FOR GENERAL SYSTEMS RESEARCH
Houston, Texas. *Contact* Brian Gaines, Department of Electrical Engineering Science, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England.

January 4–5, 1979
THE TWELFTH HAWAII INTERNATIONAL CONFERENCE ON SYSTEMS SCIENCES
Ilikae Hotel, Honolulu, Hawaii. *Contact* Ralph H. Sprague, Jr., College of Business Administration, University of Hawaii, 2404 Maile Way, Honolulu, Hawaii 96822, U.S.A.

March 12 13, 1979
ELEVENTH ANNUAL SOUTHEASTERN SYMPOSIUM ON SYSTEMS THEORY
Clemson, South Carolina. *Contact* John Bennett, Department of Electrical and Computer Engineering. Clemson University, Clemson, South Carolina 29631, U.S.A.

March 14–16, 1979
TWELFTH ANNUAL SIMULATION SYMPOSIUM
Tampa, Florida. *Contact* Joe Clema, Simulation Technology, 4124 Linden Avenue., Dayton, Ohio 45432, U.S.A.

March 1979
IFAC/IFORS SYMPOSIUM ON COMPARISON OF AUTOMATICS AND OPERATION RESEARCH TECHNIQUES APPLIED TO LARGE SYSTEMS AND CONTROL
Toulouse, France. *Contact* M. Pelegrin. CERT, B.P. 4025, 31055 Toulouse Cedex, France.

April 17–19, 1979
CONFERENCE ON THEMES IN SYSTEMS RESEARCH: PROBLEMS OF CONTEXT
Amsterdam, The Netherlands. *Contact* Gerard de Zeeuw, c/o Subfaculty for Andragology, Grote Bickersstraat 72, 1013 KS Amsterdam, The Netherlands.

June 25–29, 1979
IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY
Grignano, Italy. *Contact* L. D. DaVisson, Electrical Engineering Department, University of Maryland, College Park, Maryland 20742, U.S.A.

June 28–30, 1979
FIRST INTERNATIONAL SYMPOSIUM ON POLICY ANALYSIS AND INFORMATION SYSTEMS
Duke University, Durham, North Carolina. *Contact* Paul P. Wang, Dept. of Electrical Engineering, Duke University, Durham, N.C. 27706, U.S.A.

July 16–18, 1979
1979 SUMMER COMPUTER SIMULATION CONFERENCE
Toronto, Canada. *Contact* A. J. Schiewe, c/o The Aerospace Corp., Box 92957, Los Angeles, Ca. 90009, U.S.A.

July 16–20, 1979
SIXTH INTERNATIONAL COLLOQUIUM ON AUTOMATA, LANGUAGES AND PROGRAMMING
Technical University of Grag, Austria. *Contact* H. Maurer, Institute für Informationsverarbeitung Techn. Universität Grag, Steyrergasse 17, A-8010 Grag, Austria.

July 17–19, 1979
1979 INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS
Tokyo, Japan. *Contact* T. Fujisawa, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan

August 6–17, 1979
FIRST INTERNATIONAL SUMMERSCHOOL ON SYSTEMS METHODOLOGY
Delft, The Netherlands. *Contact* Dr. G. Broekstra, Graduate School of Management, Poortweg 6–8, . Delft, The Netherlands.

August 20–24, 1979
SIXTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE
Tokyo, Japan. *Contact* Raj Reddy, Department of Computer Science, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pa. 15213, U.S.A.

August 21–24, 1979
INTERNATIONAL MEETING OF THE SOCIETY FOR GENERAL SYSTEMS RESEARCH
London, England. *Contact* Richard Ericson, George Washington University, Lisner Hall, 2023 G. Street N. W., Washington D. C. 20006, U.S.A.

August 22–29, 1979
SIXTH INTERNATIONAL CONGRESS OF LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE
Hanover, Federal Republic of Germany. *Contact* Secretariat des Internationalen Wissenschatten Welfengarten 1, D-3000 Hanover, BRD.

September 10–14, 1979

NINTH INTERNATIONAL CYBERNETICS CONGRESS
Namur, Belgium. *Contact* J. Lemaire, Manogring Directos, IAC, Palais des Expositions, Place André Rijckmans, B-5000, Namur, Belgium.

September 24–28, 1979

IMACS CONGRESS 1979 ON SIMULATION OF SYSTEMS
Sorrento, Italy. *Contact* Centro Studi Calcolatori Ibridi, Segreteria Congresso IMACS 1979, Facolta Ingegneria, V Claudio 21, 80125 Napoli, Italy.

October 8–12, 1979

INTERNATIONAL SYMPOSIUM ON SYSTE MODELING-CONTROL
Zakopane, Poland. *Contact* Center of Electronic Computir Technical University of Lodž, UL. Piotrkowska 220, 90–3 Lodž, Poland.

Introducing
The North Holland Series in General Systems Research
Dr. George Klir, Editor

The first volume in this important new series has just been published...

Fundamentals of Measurement and Representation of Natural Systems

by **Robert Rosen,**
Killam Professor of Biomathematics at Dalhousie University, Canada

This in-depth investigation of an important approach to systems complexity examines the relationships which exist between different descriptions (i.e., between different modes of abstraction) and the manner in which different kinds of descriptions can be combined to obtain more comprehensive ones. Since most of the problems of science are concerned with relationships which exist between various systems, this book is designed to explore such relationships in a systematic way.

The author provides a comprehensive theory of observables — the basic units of all system description — and the descriptions arising from them. The theory is then applicable to any situation in which objects of interest are named or labelled by definite mapping processes, measurement in physics, pattern recognition, classification or discrimination. In the process, relationships are established between the different languages which have been used to deal with such problems in specific situations: symmetry, invariance, similarity, homeostasis. The resulting formalism is in fact a general theory of modelling: two systems model each other to the extent that one can be substituted for another in such a way that some measure of system behavior is left invariant. The study of this kind of situation provides the major thrust of the book.

The basic problem of reductionism is studied in detail. The author shows how the reductionistic hypothesis — that there is a universal mode of system analysis from which all others can be derived — is not necessarily a valid hypothesis in the natural world.

Fundamentals of Measurement and Representation of Natural Systems is a valuable addition to the personal reference library of anyone concerned with the epistemology of natural systems, or with problems of measurement recognition or classification.

1978 240 pages $22.95 0-444-00261-8

ELSEVIER NORTH-HOLLAND INC
NEW YORK 52 Vanderbilt Avenue New York, New York 10017

DIGITAL PROCESSES

VOLUME 4, NUMBER 2
SUMMER 1978

Contents

Editorial 87

A Consideration of the Use of Standard MSI Units in the Realization of Sequential Machines
D. Kidson (Bath, UK) 89

Application of Fault Folding in Test Generation for Logic Circuits
P.K. Lala and J.I. Missen (London, UK) 109

Arithmetic Codes in Residue Number Systems
F. Barsi and P. Maestrini (Pisa, Italy) 121

The Programmable Algorithm Machine (PAM) : A Proposed Family of Special-purpose Computers
M. Vineberg (San Diego, USA) 137

Short Communications

A Note on Second-order-difference Digital Differential Analysers
P.W. Baker (Kensington, Australia) 163

I^2L JK Flip-flop Circuits
C. Piguet (Neuchâtel, Switzerland) 171

Book Review

Infotech State of the Art Report on Microprocessors 177

Analyse et synthèse des systèmes logiques
D. Mange 178

© 1978, Georgi Publishing Company, CH 1813 St-Saphorin, Switzerland.

Published quarterly. Printed in Switzerland.

NOTES FOR CONTRIBUTORS

TYPESCRIPTS

Papers should be typed on good quality paper with double spacing and wide (3 cm) margins, using one side only, and submitted in triplicate to the editor: **George J. Klir, Department of Systems Science, School of Advanced Technology, State University of New York, Binghamton, New York 13901, U.S.A.**

One copy of the typescript must be the original "top" copy; Xerox or equivalent clear carbon copies are acceptable for the remaining two. Pages of each copy must be sequentially numbered. Use underlines only for emphasis and where italics are intended. Submission of a paper to this journal will be taken to imply that it represents original work not previously published, that it is not being considered elsewhere for publication, and that if accepted for publication, it will not be published elsewhere in the same form, in any language, without the consent of editor and publisher. It is a condition of the acceptance by the editor of a typescript for publication that the publishers acquire automatically the copyright in the typescript throughout the world. It will also be assumed that the author has obtained all necessary permissions to include in the paper items such as quotations, results of government sponsored research, etc. Papers are accepted only in English, and the maximum preferred length is 7500 words. Each paper requires an abstract of 150 words or less, and a set of no more than 10 index terms (key words), which, as a group, should optimally characterize the paper. For anonymity in review process, name(s) of author(s) and affiliations should appear only on a separate title page; this information will not be made available to reviewers. Enclose also a short biography and photograph of each author of a paper (not for short notes or letters), or be ready to supply these upon acceptance of the paper. For biography style, see previous issues of this journal.

FIGURES

All figures should be numbered with consecutive arabic numbers, have descriptive captions, and be mentioned in the text. Keep figures separate from the text, but indicate an approximate position for each in the margin. All figures must be of a high enough standard for direct reproduction. Line drawings should be prepared in black (India) ink on white paper or tracing cloth, with all lettering and symbols included. Alternatively, good sharp photoprints ("glossies") are acceptable. Photographs intended for halftone reproduction must be good glossy original prints, of maximum contrast. Clearly label each figure with the author's name and figure number; indicate "top" where this is not obvious. Redrawing or retouching of unusable figures will be charged to authors. Figures should be planned so that they reduce to 7.5 cm column width. The *preferred* width of submitted line drawings is 12 to 15 cm, with capital lettering 4 mm high, for reduction by one-half. Photographs for halftone reproduction should be approximately twice the desired size. A list of figure captions, with relevant figure numbers, should be typed on a separate sheet and included with the typescript.

MATHEMATICAL EXPRESSIONS

Wherever possible, mathematical expressions should be typewritten, with subscripts and superscripts clearly shown. It is helpful to identify unusual or ambiguous symbols in the margin when they first occur. To simplify typesetting please use: (1) the "exp" form of complex exponential functions; (2) fractional exponents instead of root signs; and (3) the solidus (/) to simplify fractions—e.g. $\exp x^{1/2}$. Equations must be displayed exactly as they should appear in print, and numbered in parentheses placed at the right margin. Reference to equations in the text should use the form "Eq. (5)" or simply "(5)". Do not underline mathematical symbols. The printer will set them in italics, except for obvious groups like sin and log. Any symbols which are to be left in roman (upright) type should be encircled in pencil in the typescript; bold symbols should be underlined with a wavy line.

TABLES

Tables should be typed on separate sheets, numbered consecutively with roman numerals, and have a short descriptive caption at the top. They may be placed in the typescript immediately following the page of text with which they should be printed, or they may be grouped separately; in the latter case, indicate in the text where tables are to appear. Avoid the use of vertical rules in tables. Each table should be mentioned in the text. *Extensive and/or complex tables will be reproduced as illustrations*; therefore, they should be typed carefully and in the exact format desired. Computer printouts will normally be reproduced as illustrations.

REFERENCES AND NOTES

References and notes are indicated in the text by consecutive superior arabic numbers (without parentheses). The full list should be collected and typed at the end of the paper in numerical order. Bibliographic items should include only works cited in the text and should appear in the following style:

Papers: Author's initials and last name, title of the paper in quotation marks (followed by a period), name of the journal (underlined), volume number (underlined with a wavy line), number of the issue, month and year, inclusive page numbers.

1. L. A. Zadeh, "Similarity relations and fuzzy orderings." *Information Sciences*, **3**, No. 2, April 1971, pp. 171-200.

Books: Author's initials and last name, title of the book (underlined and followed by a period), publisher, location, year, chapter or page numbers if desired.

2. G. J. Klir, *An Approach to General Systems Theory* Van Nostrand Reinhold, New York, 1969.

Edited collections: Author's initials and last name, title of the paper enclosed in quotation marks (followed by a period), name of the volume (underlined), editor's name, publisher, location, year, inclusive pages.

3. W. R. Ashby, "Systems and their informational measures." In: *Trends in General Systems Theory*, edited by G. J. Klir, John Wiley, New York, 1972, pp. 78-97.

TEXT HEADINGS

Type first-level text headings in capital letters over to the left; begin the text on the following line. Second-level headings should be typed in lower-case letters but with all main words capitalized; underline the heading and start the text on the next line. For third-level headings, only the first letter should be capital; underline, then run on the text after three typewriter spaces.

DEPOSITORY

In order to make research results quickly available to the readers, submitted typescripts may be included in a depository provided the author gives written permission for this inclusion. Abstracts of papers added to the depository are published as soon as possible in the journal and copies of their manuscripts are available for a fee, depending on the number of pages. In addition, the depository contains typescripts and other documents (computer programs, tables of empirical data, etc.) which were submitted only for the depository. Abstracts of all new additions to the depository are listed in the current issue of the journal.

PROOFS

Authors will receive page proofs (including figures) by airmail for correction, which must be returned to the printer within 48 hours. Please ensure that a full postal address is given on the first page of the typescript, so that proofs are not delayed in the post. In case of co-authorship, initial submission should indicate to whom proofs must be sent. Authors' alterations in excess of 10% of the original composition cost will be charged to authors.

Reprints: Reprints may be ordered by completing the appropriate form sent with proofs.

PAGE CHARGES

There are no page charges to individuals or to institutions.

ISSN 0308–1079
IJGS 5(1) 1–74 (1979)

International Journal of General Systems

CONTENTS Volume 5, Number 1 (1979)

APPROXIMATE, NON-DETERMINISTIC MODELLING OF BEHAVIOR SEQUENCES 1
IAN H. WITTEN

HOMOMORPHISMS IN THE THEORY OF MODELLING 13
NORMAN Y. FOO

A GENERAL SYSTEMS THEORY OF PRODUCTIVITY 17
RICHARD O. MASON

SYSTEMS AND DISTINCTIONS; DUALITY AND COMPLEMENTARITY 31
JOSEPH A. GOGUEN and FRANCISCO J. VARELA

GROWING CELLULAR SYSTEMS: GENERATION OF GRAPHS BY PARALLEL REWRITING 45
ARISTID LINDENMAYER and KAREL CULIK II

Letters to the the Editor 57

Special Book Reviews 63

Calendar 73